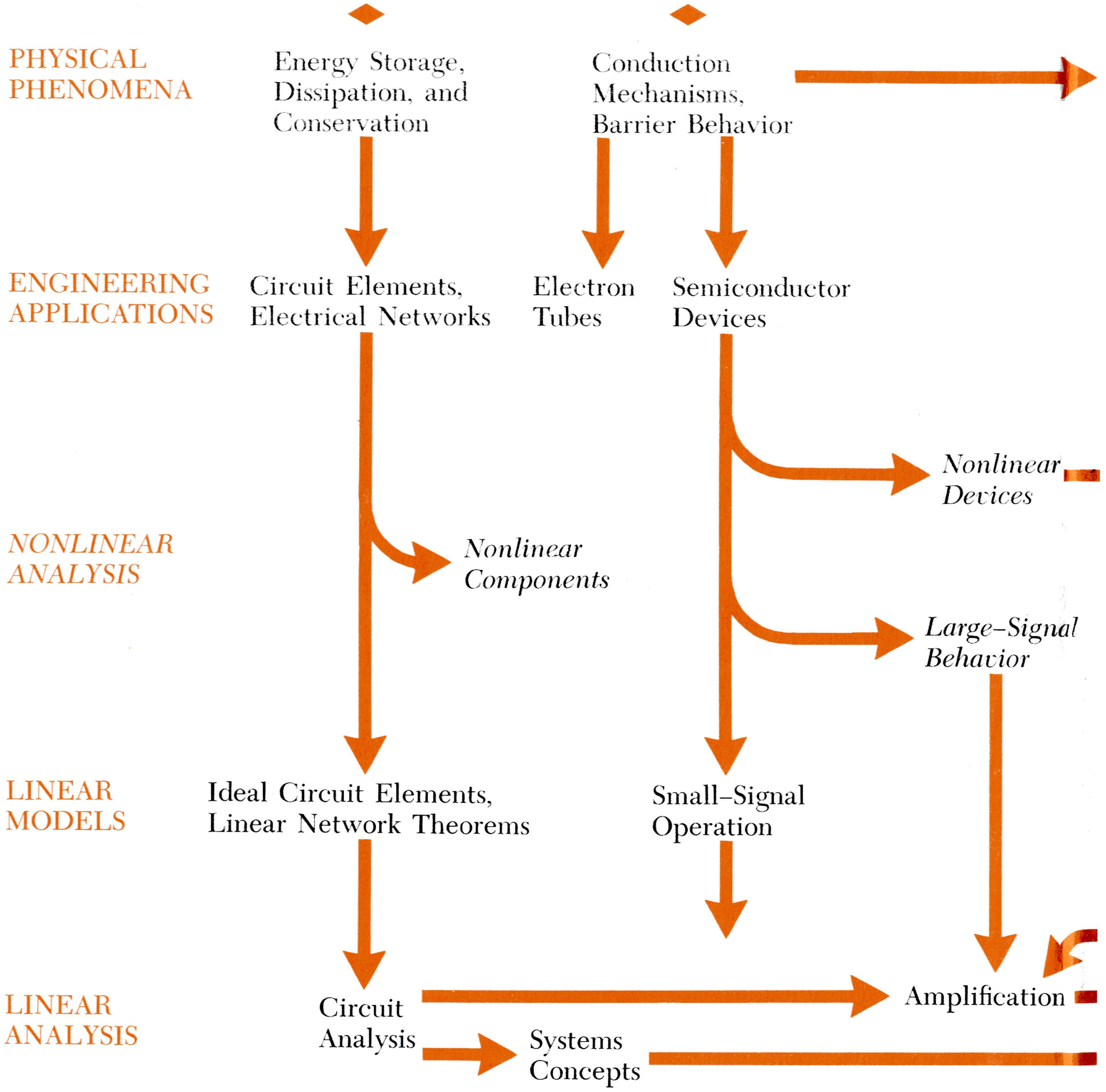
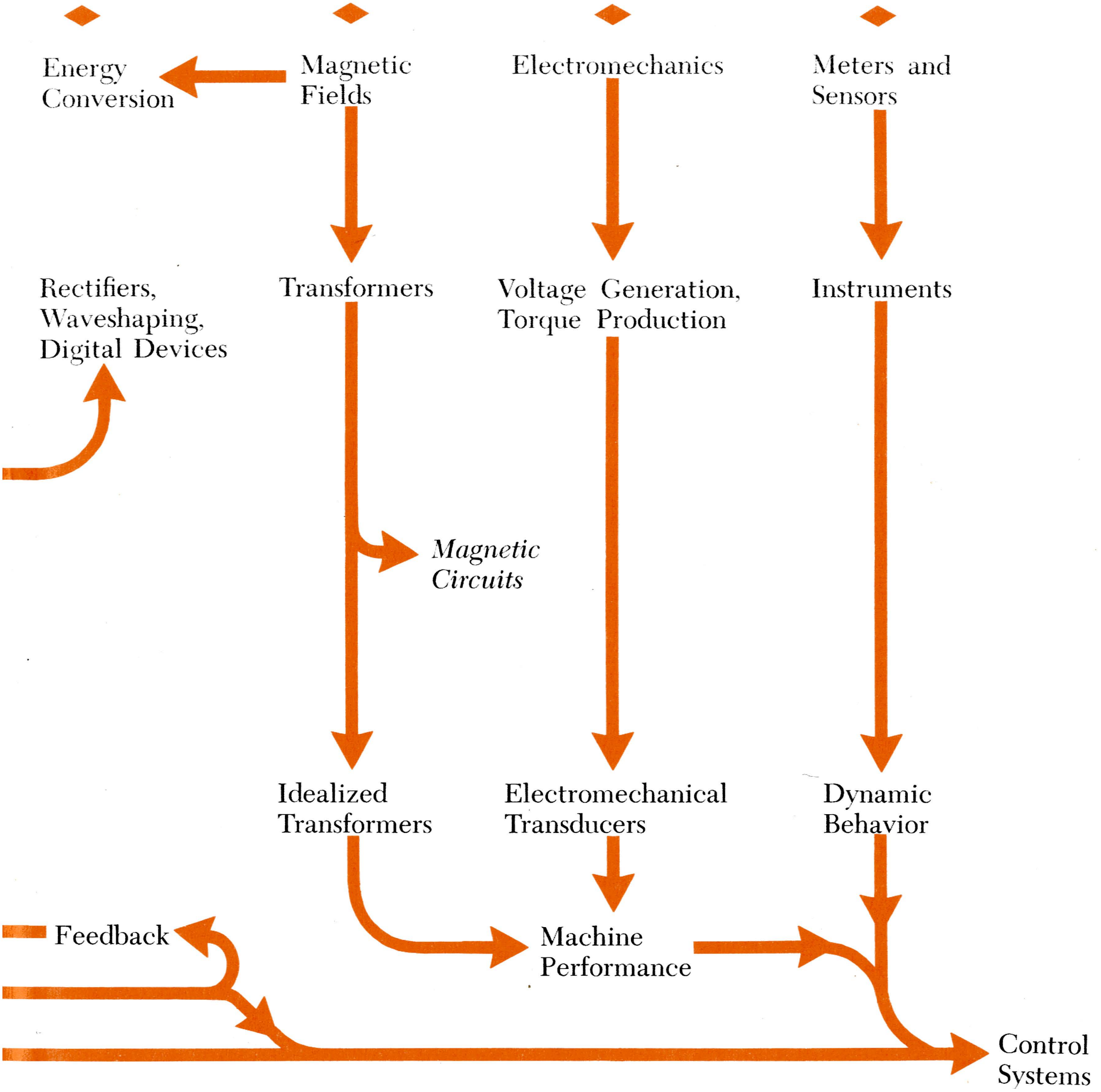
CIRCUITS DEVICES AND SYSTEMS

THIRD EDITION

RALPH J. SMITH

CONCEPT FLOW CHART


PHYSICAL PHENOMENA
Energy Storage, Dissipation, and Conservation
Conduction Mechanisms, Barrier Behavior
ENGINEERING APPLICATIONS
Circuit Elements, Electrical Networks
Electron Tubes
Semiconductor Devices
NONLINEAR ANALYSIS
Nonlinear Components
Nonlinear Devices
Large–Signal Behavior
LINEAR MODELS
Ideal Circuit Elements, Linear Network Theorems
Small–Signal Operation
LINEAR ANALYSIS
Circuit Analysis
Systems Concepts
Amplification


Energy Conversion
Magnetic Fields
Electromechanics
Meters and Sensors
Rectifiers, Waveshaping, Digital Devices
Transformers
Voltage Generation, Torque Production
Instruments
Magnetic Circuits
Idealized Transformers
Electromechanical Transducers
Dynamic Behavior
Feedback
Machine Performance
Control Systems

CIRCUITS, DEVICES, AND SYSTEMS

Also by Ralph J. Smith

Engineering as a Career, THIRD EDITION, 1969

Electronics: Circuits and Devices, 1973

CIRCUITS, DEVICES, AND SYSTEMS

A First Course in Electrical Engineering

Third Edition

Ralph J. Smith

Professor of Electrical Engineering, Stanford University

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Cover photograph by Geoffrey Gove.

Copyright © 1966, 1971, and 1976 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data

Smith, Ralph Judson.
Circuits, devices, and systems.

Includes bibliographical references and index.
1. Electrical engineering. I. Title.
TK145.S616 1976 621.3 75-29290
ISBN 0-471-80171-2

Printed in the United States of America

20 19 18 17 16 15 14 13 12

To John,
Michael, and
Kendra

PREFACE

This third edition incorporates many of the helpful suggestions of teachers and students who used the previous edition, and it reflects changes occurring in the rapidly moving field of electrical engineering. The aims of the revision have been to provide a more effective arrangement of material, a more modern emphasis in applications, and an improved set of examples and problem assignments.

The basic objectives and approach remain the same. The major changes are in content and arrangement: increased attention given to MOSFETs, ICs, SCRs, and OP AMPs; earlier and more prominent treatment of digital devices and logic circuits; and rearrangement of topics to provide an orderly progression from simple concepts to more sophisticated ideas. Minor deletions have been made to improve the balance and to make room for new material.

Objectives

This book has been designed to provide an understanding of the fundamental principles of electrical engineering, a familiarity with the viewpoint and terminology of electrical engineers, and an introduction to some modern techniques for the solution of difficult problems in all branches of engineering. The plan has been to employ modern approaches and interesting applications in developing mastery of important principles and powerful methods. The goal is the ability to predict the behavior of common circuits and devices and to design simple subsystems by specifying the critical parameters.

It is assumed that students using this book have completed basic physics courses in mechanics, electricity, and magnetism; however, the essential physical concepts are reviewed as they are introduced. It is also assumed that students have a working knowledge of differential and integral calculus. Prior knowledge of differential equations is not required; the analytical technique employed here provides a good introduction to the formulation and solution of first- and second-order linear differential equations in engineering situations.

In this first course in electrical engineering, the objectives are limited. Topics and approaches have been selected to provide the basis for further work in circuits, electronics, digital systems, electromechanics, or control systems. For electrical engineers who will later study all these in detail, a primary objective of this book is to provide a clear picture of the relationships among these subject areas. A major in another department should gain a good grasp of the fundamentals in these areas that will help him select courses for subsequent study as well as provide a solid background for such study.

Contents

The emphasis throughout is on principles rather than practice. However, the topics selected are those of "practical" importance in modern electrical engineering, and the illustrative examples are designed to appeal to students with various interests. Although a wide range of topics is included, this is more than just a survey. By limiting the discussions to the more basic aspects of each major topic, there is time to treat these aspects with sufficient rigor so that the results are meaningful, quantitative, and useful.

Part I, "Circuits," designed to provide a solid introduction to network analysis and applications, has been rearranged to improve teachability. Network theorems are introduced in Chapter 2 for early use in resistive circuits and restated for general circuits in Chapter 8. The two-wattmeter method of power measurement is added in Chapter 7 and the ideal transformer as a coupling device in Chapter 8. The new Chapter 9 introduces systems concepts such as feedback and transfer functions and provides an appropriate termination to Part I.

Part II, "Electronic Devices," has been updated to reflect recent advances, rewritten to place proper emphasis on digital devices and operational amplifiers, and reorganized to provide a smooth progression from simple concepts to sophisticated systems. Chapter 10 introduces electron motion in electric and magnetic fields, discusses cathode-ray tube operation as background for laboratory work, and briefly describes vacuum tubes. Chapters 11 and 12 are devoted to semiconductor behavior, *pn* diode operation, large-signal modeling, and application of diodes to rectification and waveshaping. Chapter 13 explains, quantitatively, the operation of field-effect and bipolar junction transistors, integrated circuits, and controlled rectifiers. This background is then employed in the analysis of digital devices and logic circuits where simple models are adequate (Chapters 14 and 15). Amplifiers are treated in four chapters starting with large-signal amplifiers where graphical analysis is satisfactory, progressing to small-signal and feedback amplifiers where more sophisticated models are required, and culminating in a thorough introduction to the operation and application of versatile operational amplifiers.

Part III, "Electromechanical Devices," presents material on energy sources and conversion (Chapter 20) that has become of great interest in our energy situation. The presentation on magnetic circuits, transformers, and electromechanics (Chapters 21 to 23) has been refined to eliminate confusion and shifted in emphasis to improve motivation. Chapters 24 and 25 on dc and ac machines replace the former long chapter on rotating machines; modern practice is reflected in the discussion of rectifiers for adjustable-frequency drive of ac machines. The emphasis is on steady-state performance, but dynamic behavior is discussed in connection with control systems (Chapter 26).

Approach

The Concept Flow Chart on the front endpaper indicates how unity has been achieved in treating diverse topics. To meet the needs and interests of engineering students, a quantitative physical explanation accompanies the description of each new device. For example, an explanation of conduction in semiconductors and potential barrier effects accompanies the description of diodes and transistors, and basic principles of electromechanics are used in deriving expressions for rotating machine performance.

To provide coherence over a broad range of topics, a few basic tools are used repeatedly. For example, the same modeling process is employed for transistors, transformers, transducers, and rotating machines. The use of simplifying assumptions to reduce a complicated system with inherent nonlinearities to a linear, relatively simple form is emphasized. Other frequently used techniques are: the pole-zero concept, the impedance function, the transfer function, the load-line approach, and block diagrams.

Organization

As in the previous editions, I have tried to help the student to learn and the instructor to teach. Again the assignments at the end of each chapter are grouped into three categories. The *Review Questions* are primarily for the student's own use in testing his understanding of the concepts and terminology introduced in that chapter. The *Exercises* vary in difficulty but, in general, they are straightforward applications of the new principles to specific situations. The *Problems* are more involved and may require extending a concept, or making simplifying assumptions, or putting ideas together in a design. Answers to the *Exercises* and solutions to the *Problems* are included in the *Instructor's Manual,* available from the publisher to teachers adopting the book for class use.

Here at Stanford, we now offer a three-term year course in circuits, electronics, and electromechanics and the book reflects that organization. Electronics is *not* a prerequisite to electromechanics. The schedule followed here and other possible arrangements for courses of different length and emphasis are included in the *Instructor's Manual,* along with a brief description of each chapter from the teacher's viewpoint. Since laboratory work is such a valuable complement to the text discussions, an appropriate laboratory course is described and sample experiment instructions are included. I will welcome correspondence on any aspect of the book or the manual.

Acknowledgments

It is a pleasure to acknowledge the assistance of many persons in this revision. Stanford colleagues Robert Helliwell and Harry Garland have given me valuable suggestions for improvement, and John R. Eliason of the General Electric Co. advised me on adjustable-speed drives. In addition, the revised manuscript was read critically by Charles A. Fowler, III, of the United States Naval Academy, Martin N. Kaplan of Drexel University, and William Parish of the University of Idaho. Although all their suggestions could not be incorporated, the book is greatly improved as a result of their comments. Among the many others who assisted in the project, I should like to thank Dan Ling, Jim Dunham, Anupam Khanna, Kai Wang, Paul Tang, and Andy Naegli for their valuable help with the exercises and problems.

The enthusiastic encouragement of John Linvill is greatly appreciated. For her ability to follow sketchy instructions and complicated text, I am indebted to Barbara Newman. For skillful collating and patient proofreading that made my work easier, I am particularly grateful to my wife Louise.

Stanford, California Ralph J. Smith

Contents

Part I
Circuits

1. ELECTRICAL QUANTITIES 3
Introduction · Definitions and laws · Circuit elements

2. CIRCUIT PRINCIPLES 29
Circuit laws · Network theorems · Nonlinear networks

3. SIGNAL WAVEFORMS 61
Exponentials · Sinusoids · Phasors · Periodic waveforms

4. NATURAL RESPONSE 87
First- and second-order systems · Impedance · Poles and zeros

5. FORCED RESPONSE 117
Responses · Phasor methods · AC analysis · Analogs and duals

6. COMPLETE RESPONSE 147
Procedure · First- and second-order circuits · Impulse response

7. STEADY-STATE AC CIRCUITS 169
Power · Frequency response · Resonance · Three-phase circuits

8. GENERAL NETWORK ANALYSIS 209
One-port network theorems · Two-port networks

9. INTRODUCTION TO SYSTEMS 231
Systems · Block diagrams · Feedback · Transfer functions

Part II
Electronic Devices

10. ELECTRON TUBES 255
Electron motion · Cathode-ray tubes · Vacuum tubes

11. SEMICONDUCTOR DIODES 287
Conduction · Doped semiconductors · Junction and special diodes

12. DIODE APPLICATIONS 311
Diode models · Rectifiers and filters · Waveshaping circuits

13. TRANSISTORS AND INTEGRATED CIRCUITS 335
Field-effect and junction transistors · Integrated circuits · Thyristors

14. DIGITAL DEVICES 363
Switching logic · Electronic switches · Memory devices

15. LOGIC CIRCUITS 387
Binary numbers · Boolean algebra · Analysis · Synthesis · Computers

16. AMPLIFIER BIAS AND LARGE-SIGNAL PERFORMANCE 409
Practical amplifiers · Biasing circuits · Power amplifiers

17. SMALL-SIGNAL MODELS 439
Large and small signals · Diodes · Field-effect transistors · Junction transistors

18. SMALL-SIGNAL AMPLIFIERS 465
RC coupled · Frequency response · Multistage · Feedback

19. OPERATIONAL AMPLIFIERS 503
Ideal op amps · Applications · Practical op amps · Analog computer

Part III
Electromechanical Devices

20. ENERGY CONVERSION PHENOMENA 535
Sources · Converters

21. MAGNETIC FIELDS AND CIRCUITS 549
Field concepts · Circuit calculations

22. TRANSFORMERS 569
AC excitation · Transformer operation · Circuit models · Performance

23. PRINCIPLES OF ELECTROMECHANICS 597
Translational transducers · Rotational · Moving-iron

24. DIRECT-CURRENT MACHINES 631
DC generators · DC motors

25. ALTERNATING CURRENT MACHINES 655
Alternators · Synchronous motors · Induction motors

26. AUTOMATIC CONTROL SYSTEMS 685
System characteristics · Dynamic response · Feedback control

27. INSTRUMENTS AND INSTRUMENTATION SYSTEMS 709
Electrical and electronic instruments · Transducers · Systems

APPENDIX 739
Physical constants · Conversion factors · μA741 Frequency compensated operational amplifier · Color code for resistors · Preferred resistance values

ANSWERS TO SELECTED EXERCISES 749

INDEX 755

CIRCUITS, DEVICES, AND SYSTEMS

PART 1
CIRCUITS

Chapter 1
Electrical Quantities

Introduction
Definitions and Laws
Circuit Elements

Energy and information are the hallmarks of civilization, and electrical engineers influence social change by improving energy conversion and information processing. In transformation of the power of falling water to electrical form in a remote hydroelectric plant, transmission at high voltage to urban load centers, and utilization in lights, motors, and industrial processes, *energy* is the significant quantity. In an observation satellite converting sensor output to digital form, performing preliminary data reduction, and sending the results back to earth, energy is merely the means for processing *information.*

In general, electrical engineers are responsible for optimizing the generation, storage, transmission, control, and conversion of electrical energy or information. But all engineers employ electrical devices and systems, and effective use of electrical components and concepts depends on mastery of the fundamentals of electrical engineering. Furthermore, techniques of device modeling and system analysis originally developed to handle purely electrical problems now find application in all branches of engineering.

Observations over the past two centuries, some random and some the result of careful experiment, have been organized, interpreted, and made available to us in the form of a relatively few basic principles. Ingenious methods have been developed for applying these principles to the analysis of existing devices and systems and in the creation of new and improved designs. This combination of fundamental principles and general techniques constitutes the subject matter of this book.

Forces and Fields

Electric charges are defined by the forces they exert on one another; experimentally, the forces are found to depend on the magnitudes of the charges, their relative positions, and their velocities. Forces due to the position of charges are called *electric* forces, and those due to the velocity of charges are called *magnetic* forces. All electric and magnetic phenomena of interest in electrical engineering can be explained in terms of the forces between charges.

In a television picture tube, the electrodes are designed to provide accelerating and deflecting electric fields to control the path of electrons in forming the picture. In a computer memory bank, small magnetic "doughnuts" become magnetized by pulses of current and then "remember" the digital information in the form of magnetic fields. In a motor, the combination of steel armature and copper winding is shaped to create an intense magnetic field through which the conductors pass as they develop torque. Fields are characteristically distributed throughout a region and must be defined in terms of two or three dimensions.

Circuits

In contrast to fields, the behavior of a *circuit* can be completely described in terms of a single dimension, the position along the path constituting the circuit. In an electric circuit, the variables of interest are the voltage and current at various points along the circuit. In circuits in which voltages and currents are constant (not changing with time), the currents are limited by resistances. In the case of a battery being charged by a generator, 100 feet of copper wire may provide a certain resistance to limit the current flow. The same effect could be obtained by 1 foot of resistance wire, or by 1 inch of resistance carbon, or by 1 mil of semiconductor.

When the dimensions of a component are unimportant and the total effect can be considered to be concentrated at a point or "lumped," the component can be represented by a *lumped parameter.* In contrast, the behavior of 100 feet of copper wire as an antenna is dependent on its dimensions and the way in which voltage and current are distributed along it; an antenna must be represented by *distributed parameters.* In this book, we are concerned with lumped parameter circuits only.

Devices

Circuits are important in guiding energy within *devices* and also to and from devices that are combined into *systems.* Electrical devices perform such functions as generation, amplification, modulation, and detection of signals. For example, at a radio broadcasting station a modulator changes the amplitude of the transmitted wave in accordance with a musical note to produce amplitude modulation. A *transducer* is a device that converts energy or information from one form to another; a microphone is a transducer that converts the acoustical energy in an input sound wave into the electrical energy of an output current.

Systems

Systems incorporate circuits and devices to accomplish desired results. A communication system includes a microphone transducer, an oscillator to provide a high-frequency carrier for efficient radiation, a modulator to superimpose the sound signal on the carrier, an antenna to radiate the electromagnetic wave into space, a receiving antenna, a detector for separating the desired signal from the carrier, various amplifiers and power supplies, and a loudspeaker to transduce electrical current into a replica of the original acoustic signal. A space-vehicle guidance system includes a transducer to convert a desired heading into an electrical signal, an error detector to compare the actual heading with the desired heading, an amplifier to magnify the difference, an actuator to energize the vernier control jets, a sensor to determine the actual heading, and a feedback loop to permit the necessary comparison.

Models

Circuits are important for another reason: Frequently it is advantageous to represent a device or an entire system by a *circuit model.* Assume that the 100 feet of copper wire mentioned previously is wound into the form of a multiturn coil and that a voltage of variable frequency is applied. If the ratio of applied voltage V to resulting current I is measured as a function of frequency, the observations will be as shown in Fig. 1.1. Over region A, the coil can be represented by a single lumped parameter (resistance R); in other words, the results obtained from the circuit of Fig. 1.1a are approximately the same as the results obtained from the actual coil of wire. Similarly, the behavior of the coil in region B can be represented by the circuit in Fig. 1.1b which contains another lumped parameter (inductance L). To represent the coil over a wide range of frequencies,

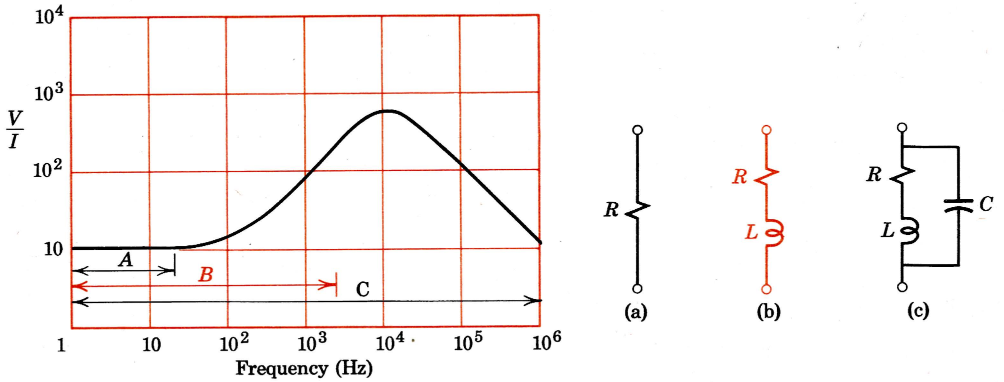
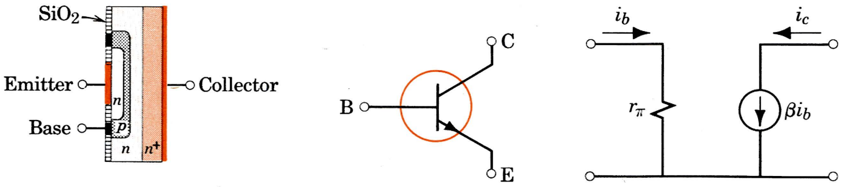


Figure 1.1 Coil characteristics and circuit models.



(a) Physical device (b) Diagram symbol (c) Circuit model

Figure 1.2 Transistor representation.

an additional parameter (capacitance C) is necessary, and the circuit model of Fig. 1.1c is used.

The technique of representing, approximately, a complicated physical device by a relatively simple model is an important part of electrical engineering. In this book we use circuit models to represent such devices as diodes and transistors (Fig. 1.2). One advantage of such a model is that it is amenable to mathematical analysis. Representing a transistor by a circuit model permits the use of well-known circuit laws in predicting the behavior of the actual transistor.

A Preview

This book has three parts. The first, Circuits, is devoted to developing the set of basic principles and powerful techniques needed to predict the behavior of a wide variety of circuits. Emphasis is placed on the fact that these principles and techniques are equally valuable in dealing with "circuits" consisting of mechanical, or thermal, or hydraulic elements. The final chapter of Part I provides an introduction to systems engineering.

In the second part, Electronics, we study the physical principles of electron motion in vacuum tubes and in semiconductors and see how these principles are employed in electronic devices, including integrated circuits. Then we examine applications of common devices in waveshaping functions, in digital circuits, and in various types of amplifiers.

The last part, Electromechanics, starts with a look at energy sources and energy conversion. Then we study the fundamentals of electromechanical energy conversion and see their application in practical devices with emphasis on rotating machines.

Although this book is primarily concerned with electrical phenomena, many of the principles, techniques, and approaches discussed are of basic importance and they find application throughout engineering.

Definitions and Laws

In beginning the study of circuits, we must first define the important circuit quantities and adopt a standard set of units, symbols, and abbreviations. Much of this material is a review of basic physics, but it deserves careful attention because it constitutes the "language" in which ideas are presented, concepts formed, and conclusions stated.

Next we take a new look at three laws based on early experiments conducted on resistors, inductors, and capacitors. Using these experimental results and modeling techniques, we invent idealized circuit components with highly desirable characteristics and then see how they behave in circuits and how they transform energy.

The International System of Units

In engineering, we must be able to describe physical phenomena quantitatively in terms that will mean the same to everyone. We need a standard set of units consistent among themselves and reproducible any place in the world. In electrical engineering, we use the SI (System International) system in which the *meter* is the unit of length, the *kilogram* the unit of mass, and the *second* the unit of time. Another basic quantity is temperature, which in the SI system is measured in *degrees Kelvin*. To define electrical quantities, an additional unit is needed; taking the *ampere* as the unit of electric current satisfies this requirement. The *candela* is needed to define illumination quantities. All quantities encountered in this book can be defined in terms of the six units displayed in Table 1-1. When data are specified in other units, they are first converted to SI units and then substituted in the applicable equations. Three conversions frequently needed are: 1 meter = 39.37 inches, 1 kilogram = 2.205 pounds (mass), and 1 newton = 0.2248 pound (force). Other useful conversion factors are given in the Appendix.

Table 1-1 Basic Quantities

Quantity	Symbol	Unit	Abbreviation
Length	l	meter	m
Mass	m	kilogram	kg
Time	t	second	s
Temperature	τ	kelvin	K
Current	i	ampere	A
Luminous intensity	I	candela	cd

Definitions

For quantitative work in circuits, we need to define the quantities displayed in Table 1-2 on p. 8. These are probably familiar from your previous study, but a brief review here may be helpful.

Force. A force of 1 newton is required to cause a mass of 1 kilogram to change its velocity at a rate of 1 meter per second per second. In this text we are concerned primarily with electric and magnetic forces.

Table 1-2 Important Derived Quantities

Quantity	Symbol	Definition	Unit	Abbreviation	(Alternate)
Force	f	push or pull	newton	N	(kg-m/s^2)
Energy	w	ability to do work	joule	J	(N-m)
Power	p	energy/unit of time	watt	W	(J/s)
Charge	q	integral of current	coulomb	C	(A-s)
Current	i	rate of flow of charge	ampere	A	(C/s)
Voltage	v	energy/unit charge	volt	V	(W/A)
Electric field strength	$\mathcal{E}$	force/unit charge	volt/meter	V/m	(N/C)
Magnetic flux density	B	force/unit charge momentum	tesla	T	(Wb/m^2)
Magnetic flux	ϕ	integral of magnetic flux density	weber	Wb	(T-m^2)

Energy. An object requiring a force of 1 newton to hold it against the force of gravity (i.e., an object "weighing" 1 newton) receives 1 joule of potential energy when it is raised 1 meter. A mass of 1 kilogram moving with a velocity of 1 meter per second possesses $\frac{1}{2}$ joule of kinetic energy.

Power. Power measures the rate at which energy is transformed. The transformation of 1 joule of energy in 1 second represents an average power of 1 watt. In general, instantaneous power is defined by

$$p = \frac{dw}{dt} \tag{1-1}$$

Charge. The integral of current with respect to time is electric charge, a concept useful in explaining physical phenomena. Charge is said to be "conservative" in that it can be neither created nor destroyed. It is said to be "quantized" because the charge on 1 electron (1.602×10^{-19} C) is the smallest amount of charge that can exist. The coulomb can be defined as the charge on 6.24×10^{18} electrons, or as the charge experiencing a force of 1 newton in an electric field of 1 volt per meter, or as the charge transferred in 1 second by a current of 1 ampere.

Current. Electric field effects are due to the presence of charges; magnetic field effects are due to the motion of charges. The current through an area A is defined by the electric charge passing through the area per unit of time. In general, the charges may be positive and negative, moving through the area in both directions. The current is the *net* rate of flow of *positive* charges, a scalar quantity. In the specific case of positive charges moving to the right and negative charges to the left, the net effect of both actions is positive charge moving to the right; the current to the right is given by the equation

$$i = +\frac{dq^+}{dt} + \frac{dq^-}{dt} \tag{1-2}$$

In a neon light, for example, positive ions moving to the right and negative electrons moving to the left contribute to the current flowing to the right. In a current of 1 ampere, charge is being transferred at the rate of 1 coulomb per second.

Voltage. The energy-transfer capability of a flow of electric charge is determined by the potential difference or voltage through which the charge moves. A charge of 1 coulomb receives or delivers an energy of 1 joule in moving through a voltage of 1 volt or, in general,

$$v = \frac{dw}{dq} \tag{1-3}$$

The function of an energy source such as an automobile battery is to add energy to the current; a 12-V battery adds twice as much energy per unit charge as a 6-V battery.

Electric Field Strength. The "field" is a convenient concept in calculating electric and magnetic forces. Around a charge we visualize a region of influence called an "electric field." The electric field strength $\boldsymbol{\mathcal{E}}$, a vector, is defined by the magnitude and direction of the force **f** on a unit positive charge in the field. In vector notation the defining equation is

$$\mathbf{f} = q\boldsymbol{\mathcal{E}} \tag{1-4}$$

where $\boldsymbol{\mathcal{E}}$ could be measured in newtons per coulomb. However, bearing in mind the definitions of energy and voltage, we note that

$$\frac{\text{force}}{\text{charge}} = \frac{\text{force} \times \text{distance}}{\text{charge} \times \text{distance}} = \frac{\text{energy}}{\text{charge} \times \text{distance}} = \frac{\text{voltage}}{\text{distance}}$$

and electric field strength in newtons per coulomb is just equal and opposite to the *voltage gradient*† or

$$\mathcal{E} = -\frac{dv}{dl} \quad \text{in V/m} \tag{1-5}$$

Magnetic Flux Density. Around a moving charge or current we visualize a region of influence called a "magnetic field." In a bar magnet the current consists of spinning electrons in the atoms of iron; the effect of this current on the spinning electrons of an unmagnetized piece of iron results in the familiar force of attraction. The intensity of the magnetic effect is determined by the magnetic flux density **B**, a vector defined by the magnitude and direction of the force **f** exerted on a charge q moving in the field with velocity **u**. In vector notation the defining equation is

---

† In the vicinity of a radio transmitter, the radiated field may have a strength of a millivolt per meter; in other words, one meter of antenna may develop a millivolt of signal voltage that can be fed into a receiver for amplification.

$$\mathbf{f} = q\mathbf{u} \times \mathbf{B} \tag{1-6}$$

A force of 1 newton is experienced by a charge of 1 coulomb moving with a velocity of 1 meter per second normal to a magnetic flux density of 1 tesla.

Magnetic Flux. Historically, magnetic fields were first described in terms of *lines of force* or *flux.* The flux lines (so called because of their similarity to flow lines in a moving fluid) are convenient abstractions that can be visualized in the familiar iron-filing patterns. Magnetic flux ϕ in webers is a total quantity obtained by integrating magnetic flux density over area **A.** The defining equation is

$$\phi = \int \mathbf{B} \cdot d\mathbf{A} \tag{1-7}$$

Because of this background, magnetic flux density is frequently considered as the derived unit and expressed in webers per square meter. In this text, however, we consider B in teslas as the primary unit.

Electrical Power and Energy

A common problem in electric circuits is to predict the power and energy transformations in terms of the expected currents and voltages. Since, by definition, $v = dw/dq$ and $i = dq/dt$, instantaneous power is

$$p = \frac{dw}{dt} = \frac{dw}{dq}\frac{dq}{dt} = vi \tag{1-8}$$

Therefore, total energy is

$$w = \int p\,dt = \int vi\,dt \tag{1-9}$$

Example 1

The "electron gun" of a cathode-ray tube provides a beam of high-velocity electrons.

(a) If the electrons are accelerated through a potential difference of 20,000 V over a distance of 4 cm (Fig. 1.3), calculate the average field strength.

(b) Calculate the average power supplied to a beam of 50 million billion electrons per second.

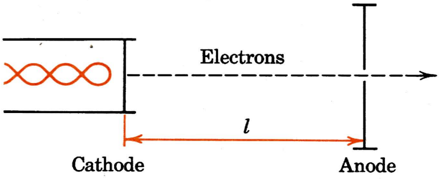


Figure 1.3 Current and power.

(a) By definition $\mathcal{E} = -dv/dl$ or

$$\mathcal{E}_{\text{av}} = \frac{\Delta v}{\Delta l} = \frac{20{,}000}{0.04} = 5 \times 10^5 \text{ V/m}$$

(b) By definition $i = dq/dt$ or

$$\begin{aligned} i &= \frac{\text{charge}}{\text{electron}} \times \frac{\text{electrons}}{\text{sec}} \\ &= 1.6 \times 10^{-19} \times 50 \times 10^6 \times 10^9 \\ &= 0.008 \text{ A} \end{aligned}$$

By Eq. 1-8 the power is

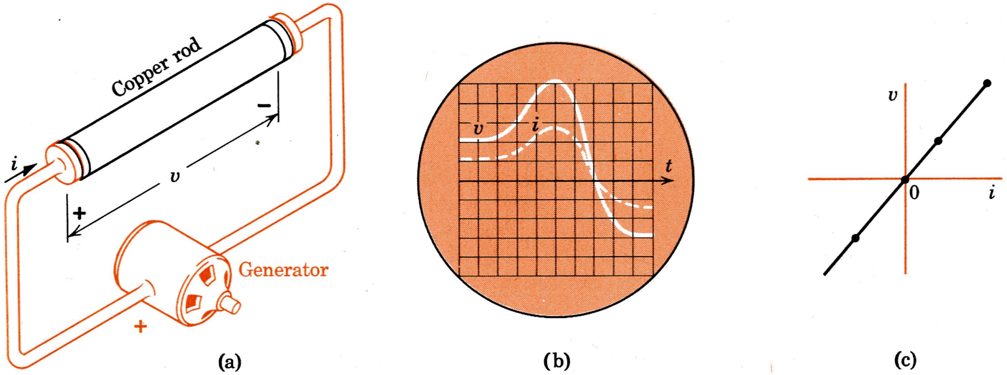
$$p = vi = 2 \times 10^4 \times 8 \times 10^{-3} = 160 \text{ W}$$

Experimental Laws

In contrast to the foregoing arbitrary definitions, three laws have been formulated (about 200 years ago) from experimentally observed facts. Our present understanding of electrical phenomena is much more sophisticated and these laws are readily derived from basic theory. At this stage in our treatment of circuits, however, it is preferable to consider these relationships as they would be revealed in experiments on real devices in any laboratory. From observations on the behavior of real devices we can derive idealized models of circuit elements and rules governing the behavior of such elements when combined in simple or complicated circuits.

Resistance

Consider an experiment (Fig. 1.4a) in which a generator is used to supply a varying current i to a copper rod. The current and resulting voltage v are observed on an oscilloscope, and voltage is plotted as a function of current (Fig. 1.4c). If all other factors (such as temperature) are held constant, the voltage is observed to be approximately proportional to the current. The experimental relation for this *resistor* can be expressed by the equation



(a) (b) (c)

Figure 1.4 Experimental determination of resistor characteristics.

$$v \cong Ri \tag{1-10}$$

where R is a constant of proportionality. Where v is in volts and i in amperes, Eq. 1-10 defines the *resistance* R in *ohms* (abbreviated Ω) and is called Ohm's law in honor of Georg Ohm, the German physicist, whose original experiments led to this simple relation. The same relation can be expressed by the equation

$$i \cong Gv \tag{1-11}$$

which defines the *conductance* G in siemens (S) or mhos (℧).

We now know that a metallic conductor such as copper contains many relatively free electrons. The application of a voltage creates an electric field that tends to accelerate these *conduction* electrons, and the resulting motion is superimposed on the random thermal motion of the electrons at, say, room temperature. Electrons are accelerated by the field, collide with copper atoms, and give up their energy. They are accelerated again, gaining energy from the electric field, collide again, and give up their energy. Superimposed on the random motion due to thermal energy, there is an average net directed motion or *drift* due to the applied electric field. The speed of drift is found to be directly proportional to the applied electric field. In a given conducting element, therefore, the rate of flow of electric charge or the current is directly proportional to the electric field, which, in turn, is directly proportional to the applied voltage. In this way, Ohm's law can be derived directly from a consideration of the conduction mechanism.

Capacitance

Now consider an experiment (Fig. 1.5a) in which a copper rod is cut and reshaped into two flat plates separated by air as an insulator. Available, but not

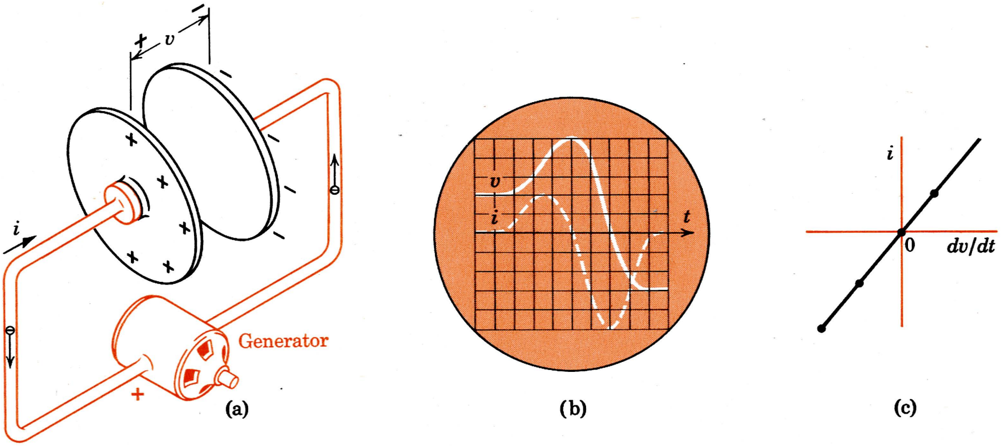


Figure 1.5 Experimental determination of capacitor characteristics.

shown, are a voltmeter, an ammeter, and an electroscope to measure voltage, current, and charge. When a voltage is applied as shown, it is observed that positive charge appears on the left-hand plate and negative on the right. If the generator is disconnected, the charge persists. Such a device that stores charge is called a *capacitor.*

As long as the voltage is constant, the charge is maintained and no current flows. If the voltage is changing with time (Fig. 1.5b), the current is observed to be approximately proportional to the *rate of change* of voltage. The experimental relation can be expressed by the equation

$$i \cong C\frac{dv}{dt} \tag{1-12}$$

where C is a constant of proportionality called *capacitance* and measured in *farads* (abbreviated F) in honor of the ingenious English experimenter, Michael Faraday.

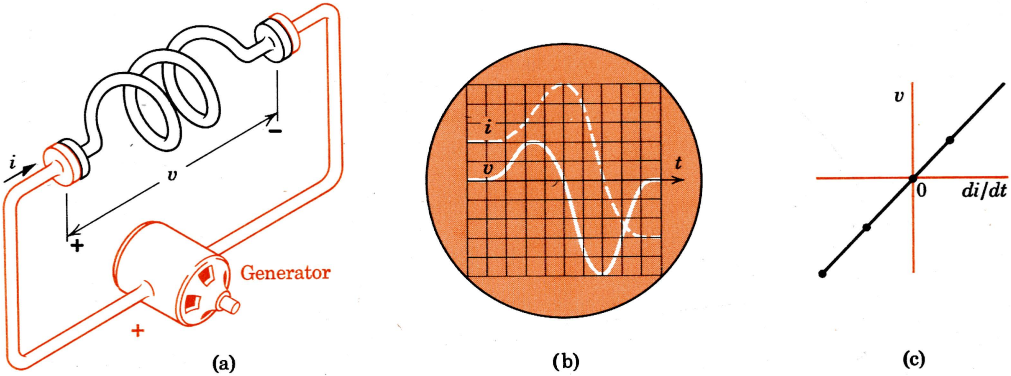
Equation 1-12 can be obtained in another way. If the charge on the capacitor is measured accurately, it is observed to be approximately proportional to the voltage applied or

$$q \cong Cv \tag{1-13}$$

where C has the same value as determined previously. Equation 1-12 can be derived from this relation since $i = dq/dt$ by definition.

Inductance

If, instead, the copper rod is drawn out and formed into the configuration of Fig. 1.6a, an entirely different result is obtained. Now the generator is used to supply current to a multiturn coil. It is observed that only a small voltage is required to maintain an appreciable steady current, but to produce a rapidly changing current a relatively large voltage is required. As indicated in Fig. 1.6c, the voltage is observed to be approximately proportional to the *rate of change* of current. The experimental relation can be expressed by the equation



(a) (b) (c)

Figure 1.6 Experimental determination of inductor characteristics.

$$v \cong L\frac{di}{dt} \tag{1-14}$$

where L is a constant of proportionality called *inductance*† and measured in *henrys* (abbreviated H) in honor of the American inventor and experimenter, Joseph Henry.

---

† More properly called "self inductance"; *mutual inductance* M is defined by the equation $v_{12} = M\, di_2/dt$ where v_{12} is the voltage induced in coil 1 due to changing current i_2 in a second coil.

This important relation can be derived by using our knowledge of magnetic fields. We know that charge in motion produces a magnetic field and the multi-turn coil is an efficient configuration for concentrating the effect of all the moving charges in the conductor. In the region in and around the coil, the magnetic flux density is high for a given current. But we also know (and this was Henry's great discovery) that a changing magnetic field *induces* a voltage in any coil that it links and the induced voltage is directly proportional to the rate of change of the magnetic field. If the magnetic field is due to the current in the coil itself, the induced voltage always tends to oppose the change in current that produces it. The voltage required of the generator is determined primarily by this voltage of *self-induction* and is approximately proportional to the time rate of change of magnetic flux density. For the coil of Fig. 1.6, magnetic flux density is proportional to current and Eq. 1-14 is confirmed.

Example 2

A current varies as a function of time as shown by the colored line in Fig. 1.7.

(a) Determine and plot the voltage produced by this current in a 30-Ω resistor.

(b) Determine and plot the voltage produced by this current in a coil of 0.5 Ω resistance and 4 H inductance.

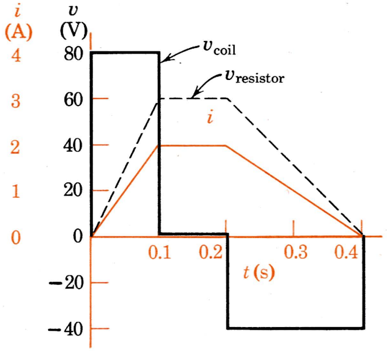


Figure 1.7 Current-voltage relations in resistance and inductance.

(a) Assuming that this is a linear resistor whose behavior is defined by $v = Ri$, the voltage is directly proportional to the current. For $i = 2$ A, $v = 30 \times 2 = 60$ V and the voltage as a function of time is as shown.

(b) For the maximum current of 2 A, the voltage due to the coil resistance is

$$v_R = Ri = 0.5 \times 2 = 1 \text{ V}$$

The voltage due to the coil inductance is:

$$0 < t < 0.1 \quad v_L = L\frac{di}{dt} = 4\frac{2}{0.1} = 80 \text{ V}$$

$$0.1 < t < 0.2 \quad v_L = L\frac{di}{dt} = 4\frac{0}{0.1} = 0$$

$$0.2 < t < 0.4 \quad v_L = L\frac{di}{dt} = 4\frac{-2}{0.2} = -40 \text{ V}$$

Later we shall learn how to combine these two effects. For many purposes, however, satisfactory results can be obtained by neglecting v_R where v_L is large and considering v_R where v_L is small. On this basis the voltage across the coil is as shown by the solid line.

Decimal Notation

To specify the *value* of a measurable quantity, we must state the *unit* and also a *number*. We shall use the SI units, but the range of numbers encountered in electrical engineering practice is so great that a special notation has been adopted for convenience. For example, the power of a public utility system may

Table 1-3 Standard Decimal Prefixes

Multiplier	Prefix	Abbreviation	Pronunciation
10^{12}	tera	T	tĕr′ ȧ
10^{9}	giga	G	jĭ′ gȧ
10^{6}	mega	M	mĕg′ ȧ
10^{3}	kilo	k	kĭl′ ō̇
10^{2}	hecto	h	hĕk′ tō̇
10^{1}	deka	da	dĕk′ ȧ
10^{-1}	deci	d	dĕs′ ĭ
10^{-2}	centi	c	sĕn′ tĭ
10^{-3}	milli	m	mĭl′ ĭ
10^{-6}	micro	μ	mī′ krō̇
10^{-9}	nano	n	năn′ ō̇
10^{-12}	pico	p	pē′ cō̇
10^{-15}	femto	f	fĕm′ tō̇
10^{-18}	atto	a	ăt′ tō̇

be millions of watts, whereas the power received from a satellite transmitter may be millionths of a watt. The engineer describes the former in *mega*watts and the latter in *micro*watts.

The notation is based on the decimal system that uses powers of 10. The standard prefixes in Table 1-3 are widely used to designate multiples and submultiples of the fundamental units. Thus 20 MW is read "20 megawatts" and is equal to 20×10^6, or 20 million watts; and 20 μW is read "20 microwatts" and is equal to 20×10^{-6}, or 20 millionths of a watt. Sometimes combinations of prefixes are used, such as kilomegahertz for a frequency of $10^3 \times 10^6$ cycles per second, but the term gigahertz (GHz) is preferred. Similarly, the term picofarads is preferred over the older designation of micromicrofarads. In Example 3 on page 16, note that all numerical quantities are first converted to SI units before being substituted in the applicable equations.

Dimensional Checking

A useful aspect of any physical quantity is how its units are related to the fundamental quantities of length, mass, time, and charge. We say that the *dimensions* of area are "length squared" or, using the conventional representation, $[L^2]$. Similarly the dimensions of velocity are $[LT^{-1}]$, force $[MLT^{-2}]$, and electric field strength $[MLT^{-2}Q^{-1}]$. The dimensions may be obtained from any equation involving the variable, neglecting dimensionless quantities such as pure numbers and derivative and integral signs. The dimensions of velocity are the same whether we work from $u \doteq \sqrt{2gh}$ or $u = dx/dt$ or $u = \int a\,dt + U_0$. (What are the dimensions of energy? Of voltage?†)

---

† Note to students: Parenthetic questions are designed to check your understanding of what has gone before or to prepare you for what is to come.

Example 3

A current varies as a function of time as shown in Fig. 1.8. Predict and plot the voltage produced by this current flowing in an initially uncharged 1-μF capacitor.

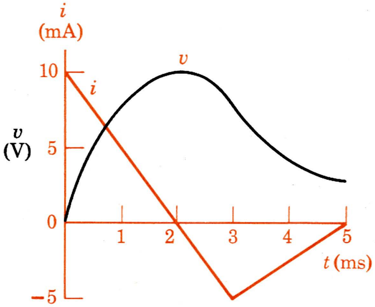


Figure 1.8 Current-voltage relations in a capacitance.

For an ideal capacitor, $i = C\,dv/dt$.

$$\therefore v = \frac{1}{C}\int_0^t i\,dt + V_0$$

For $0 < t < 3$ ms, $V_0 = 0$ and

$$i = 10 \times 10^{-3} - (15 \times 10^{-3}/3 \times 10^{-3})t$$
$$= 0.01 - 5t$$
$$\therefore v = \frac{1}{10^{-6}}\int_0^t (0.01 - 5t)\,dt + 0$$
$$= 10^6(0.01t - 2.5t^2), \text{ a parabola}$$

At $t = 2$ ms, for example,

$$v = 10^6(0.01 \times 2 \times 10^{-3} - 2.5 \times 4 \times 10^{-6})$$
$$= 20 - 10 = 10 \text{ V}$$

A similar calculation for $3 < t < 5$ ms (with $V_3 = 7.5$ V) yields the curve shown.

Dimensional checking is based on the fact that an analytical equation must be dimensionally homogeneous; every term must have the same dimensions. If an equation has been derived for velocity, every term must have the dimensions $[\mathrm{LT}^{-1}]$; if not, a mistake has been made. (See Example 4 on the next page.)

Circuit Elements

An electric circuit consists of a set of interconnected components. Components carrying the same current are said to be connected in *series*. A simple series circuit might include a battery, a switch, a lamp, and the connecting wires. Components subjected to the same voltage are said to be connected in *parallel*. The various lights and appliances in a residence constitute a parallel circuit. The terms *circuit* and *network* are sometimes used interchangeably, but usually network implies a more complicated interconnection such as the communication network provided by the telephone company. There is also an implication of generality when we refer to network analysis or network theorems.

Circuit Components

From our standpoint, the distinguishing feature of a circuit component is that its behavior is described in terms of a voltage-current relation at the terminals of the component. This relation, called the "*v*-*i* characteristic," may be obtained

Example 4

(a) For a mechanical system (which will be studied in Chapter 4) the following equation has been derived:

$$\tfrac{1}{2}M\frac{du}{dt} + Du + K\int u\,dt = f$$

where u is velocity in m/s and K is the spring compliance in m/N. What must be the units of D?

(b) Check the above equation dimensionally.

Since the right-hand side of the equation is force with the dimensions $[\mathrm{MLT^{-2}}]$, every term on the left must have the same dimensions. Therefore,

$$[\mathrm{D}][\mathrm{LT^{-1}}] = [\mathrm{MLT^{-2}}]$$

or

$$[\mathrm{D}] = \frac{[\mathrm{MLT^{-2}}]}{[\mathrm{LT^{-1}}]} = [\mathrm{MT^{-1}}]$$

and the SI units of D are kilograms per second.

The dimensions of the first term are $[\mathrm{M}][\mathrm{LT^{-1}}]/[\mathrm{T}] = [\mathrm{MLT^{-2}}]$, which is correct for force. Notice that there is no confirmation of the validity of the factor $\frac{1}{2}$; it looks suspicious and should be checked by some other method.

The units of K are m/N; therefore, the dimensions of the last term are

$$[\mathrm{L/MLT^{-2}}][\mathrm{LT^{-1}}][\mathrm{T}] = [\mathrm{M^{-1}LT^{2}}] \neq [\mathrm{MLT^{-2}}]$$

The dimensions are not those of force and, therefore, the last term is not right. (Would inverting K help?)

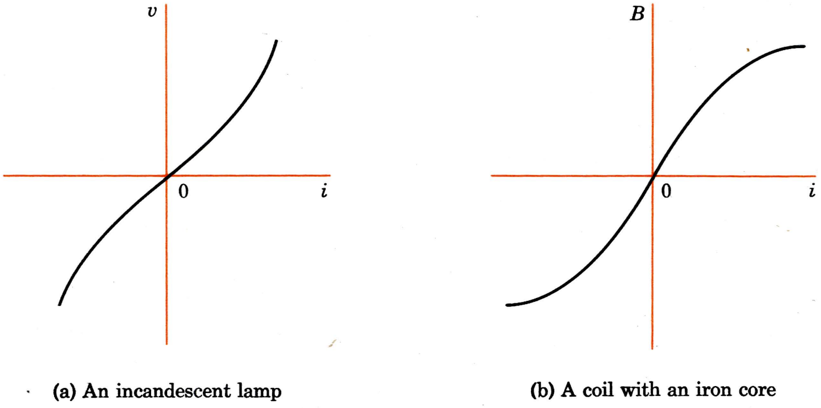
analytically by using field theory in which the geometry of the associated electric or magnetic field is important, or it may be obtained experimentally by using measurements at the terminals. Once the v-i characteristic is known, the behavior of the component in combination with other components can be determined by the powerful methods of *circuit theory*, the central topic of this part of the book.

Using hypothetical experiments and applying elementary field concepts, we obtained v-i characteristics for a resistor, a capacitor, and an inductor. In writing mathematical expressions for the v-i characteristics (Eqs. 1-10, 1-12, and 1-14) we neglected the distributed nature of the fields involved and we assumed linearity. But we know that no real physical device is exactly linear. For example, the resistance of the incandescent lamp of Fig. 1.9a increases rapidly with current, and the inductance of the coil of Fig. 1.9b is greatly affected by the value of average current.† Therefore we shall invent three ideal circuit components, or *models*, which are linear. (Figure 1.9 is on page 18.)

There are good reasons for doing this. Some real components can, within limits, be represented by such lumped linear models in many problems. Other real components can be represented accurately by combinations of two or more ideal

---

† Since v is proportional to dB/dt and, in general, B is a nonlinear function of i.



(a) An incandescent lamp (b) A coil with an iron core

Figure 1.9 Characteristics of real circuit components.

components. Also, by using the mathematical methods applicable to linear models we can derive general results of great value in predicting the behavior of complicated devices and systems of devices.

Circuit Element Definitions

The three linear models are shown symbolically in Fig. 1.10 along with their defining equations.† As you may recall, a real coil or inductor exhibits the

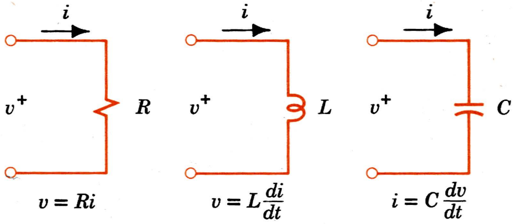


Figure 1.10 Circuit elements.

property of inductance; but every real inductor also exhibits the property of resistance. The corresponding linear model is called an "ideal inductor" or, since it exhibits only a single property, it is called an "inductance element," or simply an "inductance." The three linear models—resistance, inductance, and capacitance—are referred to as *circuit elements;* with these elements as building blocks, an infinite number of circuits can be devised.

---

† The plus sign above the "v" indicates that the upper terminal is the positive reference direction of potential. The arrow under the "i" indicates that flow to the right is the positive reference direction of current. These conventions are discussed in detail in Chapter 2.

Energy Storage in Linear Elements

Valuable insights into the behavior of real circuit components can be obtained by considering the energy transformations that occur in the corresponding linear models. Recalling that energy $w = \int vi\,dt$, we see that, once the v-i characteristic is defined, the energy storage or dissipation property is determined.

Inductance. Where $v = L\,di/dt$ and $i = 0$ at $t = 0$,

$$w_L = \int_0^T L\frac{di}{dt}i\,dt = \int_0^I Li\,di = \tfrac{1}{2}LI^2 \qquad (1\text{-}15)$$

The total energy input to an inductance is directly proportional to the square of the final current. The constant of proportionality is $L/2$; in fact, this expression for energy could have been used to define inductance.

> *Inductance is a measure of the ability of a device to store energy in the form of moving charge or in the form of a magnetic field.*

The equation reveals that the energy is stored rather than dissipated. If the current is increased from zero to some finite value and then decreased to zero, the upper limit of the integration becomes zero and the net energy input is zero; the energy input was stored in the field and then returned to the circuit.

Capacitance. Where $i = C\,dv/dt$ and $v = 0$ at $t = 0$,

$$w_C = \int_0^T vC\frac{dv}{dt}dt = \int_0^V Cv\,dv = \tfrac{1}{2}CV^2 \qquad (1\text{-}16)$$

Can *you* interpret this equation? In words, the total energy input to a capacitance is directly proportional to the square of the final voltage. The constant of proportionality is $C/2$.

> *Capacitance is a measure of the ability of a device to store energy in the form of separated charge or in the form of an electric field.*

The energy is stored rather than dissipated; if the final voltage is zero, the energy stored in the field is returned to the circuit.

These expressions for stored energy remind us of the expressions for kinetic energy of a moving mass and potential energy of a stretched spring. In such mechanical systems, displacement $dx = u\,dt$ and $w = \int f\,dx = \int fu\,dt$.

Mass. Where $f = M\,du/dt$ and $u = 0$ at $t = 0$,

$$w_M = \int_0^T M\frac{du}{dt}u\,dt = \int_0^U Mu\,du = \tfrac{1}{2}MU^2 \qquad (1\text{-}17)$$

As expected, this equation indicates that the mass M of a body is a measure of its ability to store kinetic energy.

Compliance. Where $x = Kf$ and $f = 0$ at $t = 0$,

$$w_K = \int_0^F f\,d(Kf) = \int_0^F Kf\,df = \tfrac{1}{2}KF^2 \tag{1-18}$$

In a similar way, the spring compliance K is a measure of the spring's ability to store potential energy.

Energy Dissipation in Linear Elements

Each of these elements (and there are others in thermal and chemical systems) has the ability to store energy and then return it to the circuit or system. When a similar analysis is made of an electrical resistance, the results are quite different.

Resistance. Where $v = Ri$ and $i = 0$ at $t = 0$,

$$w_R = \int_0^T Ri\,i\,dt = \int_0^T Ri^2\,dt \tag{1-19}$$

To evaluate the total energy supplied, we must know current i as a function of time t. For the special case where the current is constant or $i = I$,

$$w_R = \int_0^T RI^2\,dt = RI^2T \tag{1-20}$$

In general, for finite values of i (either positive or negative) and finite values of t, the energy supplied to the resistor is finite and positive. There is no possibility of controlling the current in such a way as to return any energy to the circuit; the energy has been *dissipated.* In a real resistor, the dissipated energy appears in the form of heat; in describing the behavior of the linear model we say that the energy has been dissipated in an *irreversible transformation,* irreversible because there is no way of heating an ordinary resistor and obtaining electric energy.

The *rate* of dissipation of energy or power is a useful characteristic of resistive elements. By definition, $p = dw/dt$; therefore,

$$p_R = \frac{dw_R}{dt} = Ri^2 \tag{1-21}$$

The power dissipation in a resistance is directly proportional to the square of the current. The constant of proportionality is R and this expression for power is frequently used to define resistance.

Resistance is a measure of the ability of a device to dissipate power irreversibly.

Friction. At this point an alert student would wonder if there is in mechanical systems a corresponding dissipative element. There is and it is called *frictional* *resistance.* The shock absorber is a friction device designed to dissipate the energy received by an automobile when it goes over a bump in the road. The shock absorber, or "dash pot," contains a viscous fluid in a piston-cylinder arrangement. Relative motion of the body and chassis forces the fluid through a small hole and the kinetic energy is converted to heat in an irreversible process.

Over limited ranges, the friction force developed is approximately proportional to velocity or

$$f = Du \tag{1-22}$$

where D is the frictional resistance in units of newtons/(meters per sec) or newton-seconds/meter. In the linear model of a dash pot where $f = Du$ and $u = 0$ at $t = 0$,

$$w_D = \int_0^X f\,dx = \int_0^T Du\,u\,dt = \int_0^T Du^2\,dt \tag{1-23}$$

and the power dissipation in friction is

$$p_D = \frac{dw}{dt} = Du^2 \tag{1-24}$$

Continuity of Stored Energy

It is common knowledge that a massive body tends to oppose rapid changes in velocity; we say that such a body has *inertia.* In solving problems in dynamics we take advantage of the fact that the velocity of a mass cannot be changed instantaneously; for example, the velocity just after a force is applied must be the same as that just before the force is applied. Similar conclusions can be drawn about electric circuit quantities by applying the fundamental principle that in any physical system the energy must be a continuous function of time.

Since power is the time rate of change of energy ($p = dw/dt$), an instantaneous change in energy would require an infinite power. But the existence of an infinite power is contrary to our concept of a physical system, and we require that the energy stored in any element of a system, real or ideal, be a continuous function of time. Recalling that the energy of a moving mass is $\frac{1}{2}Mu^2$, we conclude that the velocity u cannot change instantaneously.

Following the same reasoning, we note that the energy stored in an inductance is $\frac{1}{2}Li^2$ and, therefore, *the current in an inductance cannot change instantaneously.* Since the energy stored in a capacitance is $\frac{1}{2}Cv^2$, *the voltage across a capacitance cannot change instantaneously.* These concepts are particularly useful in predicting the behavior of a circuit just after an abrupt change, such as closing a switch. Note that in considering energy changes there is no limitation on the rapidity with which inductance voltage or capacitance current can change. Also, since resistance does not store energy, there is no limitation on the rapidity with which resistance voltage and current changes can occur.

Example 5

The ignition coil of an automobile can be represented by a series combination of inductance L and resistance R (Fig. 1.11). A source of voltage V, the battery, is applied through contacts or "points" represented by switch S.

(a) With the switch open, the current i is zero. What is the current just after the switch is closed?

(b) After a finite time the current has reached a value of 1 A and the switch is suddenly opened. What is the current just after the switch is opened?

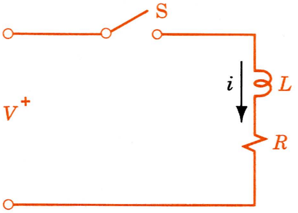


Figure 1.11 Ignition coil.

Since the current in the inductance cannot change instantaneously, the current just after closing the switch must be zero (but increasing rapidly for proper operation of the ignition system).

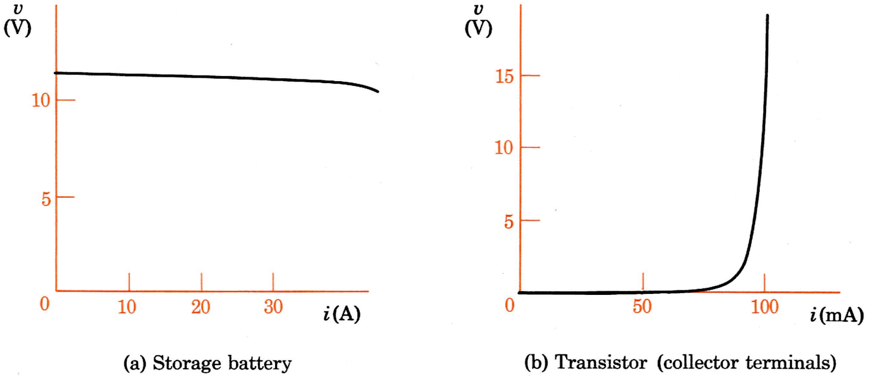
The current just after the switch is opened is still 1 A because it takes time to dissipate the energy stored in the inductance ($\frac{1}{2}Li^2$). Some of the energy goes to ionize the air between the switch contacts and an arc is formed; the points are "burned."

In practice, a "condenser" (a capacitor) is connected across the switch to minimize arcing by absorbing energy from the coil. The role of the condenser will become more clear in Chapter 4.

Energy Sources and Reversible Transformations

Energy can be stored in inductance or capacitance and dissipated in resistance. These are called *passive* circuit elements in contrast to *active* elements that can serve as sources of electrical energy. The many different forms of electrical "generators" include the rotating dynamo, the chemical storage battery, the solar cell, the fuel cell, and the thermocouple. In each of these real devices the electrical energy is "generated" by the conversion of some other form of energy. The conversion represents a *reversible transformation* in that the conversion process can go either way, in contrast to the situation in a resistor. To illustrate: in a dynamo operating as a "generator," mechanical energy is converted into electrical energy; but electrical input to a typical dynamo will cause it to operate as a "motor" producing mechanical energy.

Although every generator supplies both voltage and current ($w = \int vi\, dt$), it is desirable to distinguish between two general classes of devices. If the voltage output is relatively independent of the circuit to which it is connected (as in



(a) Storage battery (b) Transistor (collector terminals)

Figure 1.12 Characteristics of real voltage and current generators.

the battery of Fig. 1.12a), the device is called a "voltage generator." If over wide ranges the output current tends to be independent of the connected circuit (as in the transistor of Fig. 1.12b), it is treated as a "current generator."

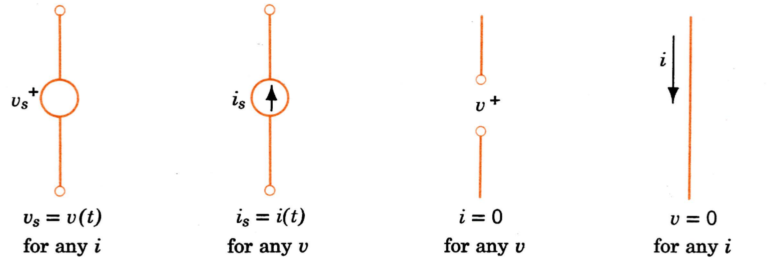
An ideal voltage generator in which the output voltage is completely independent of the current is called a "pure voltage source," or simply a "voltage source." The output voltage is specified, usually as a function of time, and is unaffected by changes in the circuit to which it is connected. At low currents, an automobile battery supplies a nearly constant voltage and may be represented by a 12-V voltage source. At high currents, however, the output voltage is appreciably less than 12 V.† The difference is due to the fact that the energy-conversion process is not perfect; part of the chemical energy is converted to heat in an irreversible process. This result suggests the possibility of representing a real electrical generator by a linear model consisting of a pure voltage source and a resistance, the resistance accounting for the irreversible energy transformations.

The circuit symbol and the defining equation for a voltage source are shown in Fig. 1.13a. The corresponding ideal current generator or "current source" is shown in Fig. 1.13b. Note that here the output current is specified, usually as a function of time, and is independent of the voltage across the source. As we shall see, a combination of a current source and a resistance also may be used to represent a real generator in predicting the behavior of real circuits. There is a finite limit to the power supplied by such a combination, whereas a pure voltage or current source can supply, or absorb, an unlimited amount of power.

Two additional circuit elements need to be defined before we begin circuit

---

† How is this fact evidenced when an automobile is started with its headlights on?



(a) Voltage source (b) Current source (c) Open circuit (d) Short circuit

Figure 1.13 Other circuit elements.

analysis. The *open circuit* shown symbolically in Fig. 1.13c is characterized by finite voltage and zero current; the *short circuit* is characterized by finite current and zero voltage. (To what real components do open and short circuits correspond?) While other circuit elements are conceivable (and, in fact, will be desirable later on), the passive elements R, L, and C, the active sources $v(t)$ and $i(t)$, and the open and short circuits constitute a set that will be sufficient to describe a large number of useful circuits.

SUMMARY

◆ The important electrical quantities, their definitions, and their SI units are summarized in Table 1-2 on p. 8.
In electrical terms, power and energy are expressed by:

$$p = \frac{dw}{dt} = vi \quad \text{and} \quad w = \int p\,dt = \int vi\,dt$$

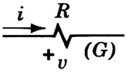
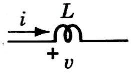
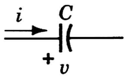
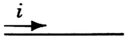
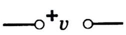
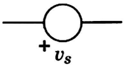
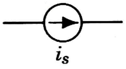
◆ Derived expressions can be checked dimensionally since an analytical equation must be dimensionally homogeneous.

◆ The approximate behavior of real circuit components is used to define the five passive and two active circuit elements displayed in Table 1-4.
Voltage and current sources represent reversible energy transformations.

◆ Resistance is a measure of the ability of a device to dissipate energy irreversibly ($p_R = Ri^2$).
Inductance is a measure of the ability of a device to store energy in a magnetic field ($w_L = \frac{1}{2}Li^2$).
Capacitance is a measure of the ability of a device to store energy in an electric field ($w_C = \frac{1}{2}Cv^2$).

◆ Energy stored in an element must be a continuous function of time.
The current in an inductance cannot change instantaneously.
The voltage on a capacitance cannot change instantaneously.

Table 1-4 Electrical Circuit Elements

Element	Unit	Symbol	Characteristic
Resistance (Conductance)	ohm (mho)		$v = Ri$ $(i = Gv)$
Inductance	henry		$v = L\frac{di}{dt} \quad i = \frac{1}{L}\int_0^t v\,dt + I_0$
Capacitance	farad		$i = C\frac{dv}{dt} \quad v = \frac{1}{C}\int_0^t i\,dt + V_0$
Short circuit			$v = 0$ for any i
Open circuit			$i = 0$ for any v
Voltage source	volt		$v = v_s$ for any i
Current source	ampere		$i = i_s$ for any v

REVIEW QUESTIONS

(These questions are primarily for use by the student in checking his or her familiarity with the ideas in the text.)

1. Without referring to the text, list with their symbols and units the six quantities basic to the SI system.
2. Without referring to the text, define "force," "energy," and "power," and give their symbols and units.
3. Without referring to the text, define "current," "voltage," "electric field strength," and "magnetic flux density," and give their symbols and units.
4. Some automobile manufacturers are considering going to an 18-V battery. What specific advantages would this have?
5. Explain the minus sign in Eq. 1-5.
6. Given two points a and b in a field of known strength $\mathcal{E}$, how could voltage v_{ab} be calculated?
7. Explain the difference between an "electric signal" that moves along a wire at nearly the speed of light and a "electric charge" that drifts along at only a fraction of 1 m/sec.
8. Define "resistance," "inductance," and "capacitance" in terms of voltage-current characteristics and also in terms of energy transformation characteristics.
9. Draw a curve representing current as a function of time and sketch on the same graph curves of the corresponding voltage across inductance and capacitance.
10. Draw a curve representing voltage as a function of time and sketch on the same graph curves of the corresponding current in an inductance and a capacitance.

11. How can a current flow "through" a capacitor containing an insulator?

12. Explain the basis for dimensional checking.

13. Once the effect of an electric field has been represented by a circuit element, is the geometry of the field significant?

14. List three physical components not mentioned in the text which have the ability to store energy.

15. The concept of continuity of stored energy places what restrictions on the circuit behavior of inductances and capacitances?

16. Give three examples of reversible transformations involving electrical energy.

17. Explain, in terms of source characteristics, the dimming of the house lights when the refrigerator motor starts.

18. What real circuit component corresponds to a short circuit on a circuit diagram? To an open circuit?

19. Distinguish between an "inductor" and an "inductance;" give an example of each.

EXERCISES

(These exercises are intended to be straightforward applications of the concepts of this chapter.)

1. A man weighing 225 lb runs up a 15-ft (vertical distance) flight of stairs in 3 s.
   (a) Express in newtons the force such a man would exert on a bathroom scale and in kilograms the "reading" of the scale.
   (b) Express in joules and in kilowatt-hours net work done in climbing the stairs.
   (c) Express in watts the average power expended in climbing the stairs.

2. A glass tube with a cross-sectional area of 10^{-2}m^2 contains an ionized gas. The densities are 10^{13} positive ions/m^3 and 10^{11} free electrons/m^3. Under the influence of an applied voltage, the positive ions are moving with an average axial velocity of 5×10^3 m/s. At that same point the axial velocity of the electrons is 2000 times as great. Calculate the electric current.

3. The potential energy lost by an electron in moving from point 1 to point 2 in a circuit is 8×10^{-15} J. Calculate the potential of point 2 with respect to point 1, and predict the work required to move a charge of 5×10^{-6} C from point 1 to point 2 in 0.1 s.

4. A voltage of 2000 V exists across the 1-cm insulating space between two parallel conducting plates. An electron of mass $m = 9.1 \times 10^{-31}$ kg is introduced into the space.
   (a) Calculate the electric field strength in the space.
   (b) Calculate the force on the electron and the resulting acceleration.

5. A 12-V automobile light is rated at 30 W.
   (a) What total charge flows through the filament in 1 min?
   (b) How many electrons does this represent?
   (c) What is the cost of 24 hr's operation at 6 cents per kilowatt-hour?
   (d) What is the resistance of the hot filament?
   (e) How many 100-μF capacitors would be required to store at 12 V the charge for 1 min's operation?

6. Ammeter AM and voltmeter VM, connected as shown in Fig. 1.14, measure instantaneous current and voltage. An ammeter reads "upscale" or positive when current flows into the meter at the + terminal. Is power flowing *into* or *out* of device D when:
   (a) AM reads positive and VM reads negative?
   (b) AM reads negative and VM reads positive?

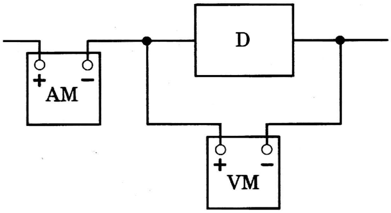


Figure 1.14

(c) AM reads negative and VM reads negative?

7. In a TV picture tube, the accelerating voltage is 25 kV.
   (a) Express in joules the energy gained by an electron moving through this potential.
   (b) If the power delivered to the fluorescent screen by the high-velocity electrons is to be 2 W, calculate the necessary beam current.
8. Show that the following have the dimensions of energy:
   (a) $\frac{1}{2}CV^2$
   (b) $\frac{1}{2}Q^2/C$
   (c) $\frac{1}{2}LI^2$
   (d) I^2RT
9. Show that the following mathematical models describe a resistance, capacitance, or inductance.
   (a) $v = kq$
   (b) $v = k\,dq/dt$
   (c) $v = k\,d^2q/dt^2$
10. The current in Fig. 1.15 is $i(t) = 5t$ A where t is in seconds. For time $t = 2$ ms, predict the voltages across the 200-Ω resistor and across the 3-H inductor.

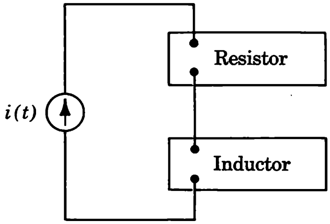


Figure 1.15

11. The voltage in Fig. 1.16 is $v(t) = 200t$ V where t is in seconds. For time $t = 50$ ms, predict the current through the 4-Ω resistor and through the 300-μF capacitor.

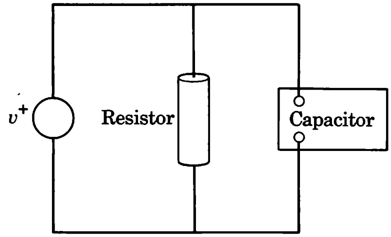


Figure 1.16

12. The current shown in Fig. 1.17 flows through a 5-H inductor and an initially uncharged 2-μF capacitor. Plot to scale graphs of v_L and v_C.

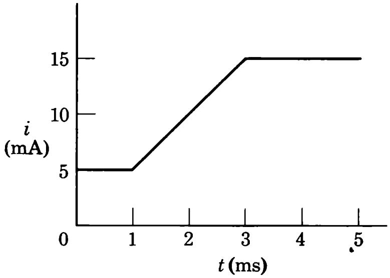


Figure 1.17

13. The voltage shown in Fig. 1.18 is applied to a 10-μF capacitor and a 0.5-H inductor (no initial current). Plot to scale graphs of i_C and i_L.

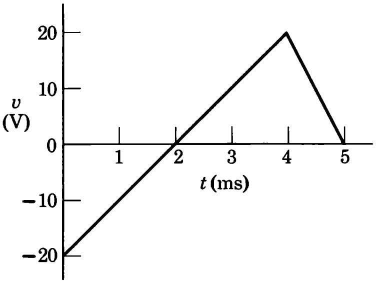


Figure 1.18

14. (a) Calculate the energy stored in an inductance $L = 10$ H at time $t = 2$ ms by the waveform of Fig. 1.17.
    (b) Calculate the energy stored in a capacitance $C = 100$ μF at time $t = 3$ ms by the waveform of Fig. 1.18.
    (c) Explain the difficulty in performing calculation (a) on Fig. 1.18 and calculation (b) on Fig. 1.17.
15. Plot as a function of time the power to the inductor of Exercise 12 and the energy stored in the inductor.
16. Plot as a function of time the power to the capacitor of Exercise 13 and the energy stored in the capacitor.

17. The current source in the circuit of Fig. 1.15 provides an output defined by $i = 5t$ A where t is in seconds. For time $t = 2$ s, calculate:
  - (a) The power dissipated in the 15-Ω resistor.
  - (b) The power supplied to the 10-H inductor.
  - (c) The power delivered by the source.
  - (d) The total energy dissipated in the resistor.
  - (e) The total energy stored in the inductor.

18. The voltage source in the circuit of Fig. 1.16 provides an output defined by $v = 200t$ V where t is in seconds. For time $t = 0.5$ s, calculate:
  - (a) The power dissipated in the 200-Ω resistor.
  - (b) The power supplied to the 1000-μF capacitor.
  - (c) The power delivered by the source.
  - (d) The total energy dissipated in the resistor.
  - (e) The total energy stored in the capacitor.

19. In Fig. 1.19, capacitor C is "charged up" by closing switch S_1. After equilibrium is reached, switch S_2 is closed.
  - (a) Just *before* closing S_2, what is the voltage across C? The current in inductor L?
  - (b) Just *after* closing S_2, what is the voltage across C? The current i_L? The current i_2?

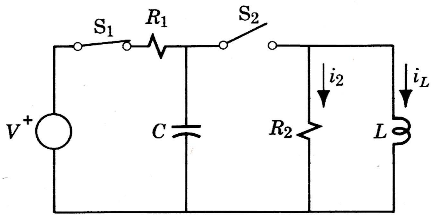


Figure 1.19

PROBLEMS

(These problems require extending the basic concepts of this chapter to new situations.)

1. Would it be physically possible to have a passive circuit component for which the voltage was proportional to the second derivative of current? If a current $i = I_m \sin t$ flowed in such a component, what power would be absorbed?

2. Laboratory tests on a coil yield the following data: a change in current from +10 mA to −10 mA in 50 μs results in an average voltage of 2 V; a steady current of 10 mA requires a power input of 0.2 W. Devise a linear circuit model to represent the coil.

3. A "Q-switching" carbon dioxide laser capable of producing 50 W of continuous-wave power will produce 50 kW of pulsed power in bursts 150 ns long at a rate of 400 bursts per second. Calculate the average power output in the pulsed mode and the energy per burst.

4. As we shall see (in Chapter 12), the circuit shown in Fig. 5.24 can be used to smooth out or "filter" variations in input current i so that a fairly constant voltage appears across load resistor R. For good filtering, the capacitor should be able to store 10 times as much energy as is dissipated in one cycle by R. If $R = 10$ kΩ and variations in i occur 60 times per sec., specify the value of C.

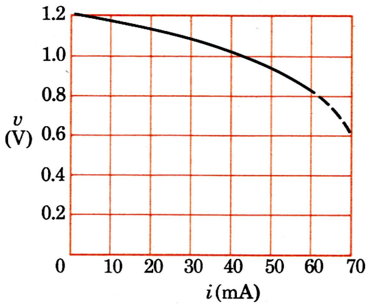


Figure 1.20

5. Devise a linear model incorporating a current source to represent the solar battery whose output characteristics are shown in Fig. 1.20. Specify numerical values valid for currents up to 40 mA.

6. Check the following equations dimensionally, indicate any errors, and suggest possible corrections:
  - (a) $v = RI\,e^{-t/LR} \sin(Rt/2L)$
  - (b) $t = -\dfrac{2C}{G} + (LC - 4C^2/G)^{1/2}$

(Note that you need not understand the phenomena to find errors in the results.)

Chapter 2
Circuit Principles

Circuit Laws
Network Theorems
Nonlinear Networks

The emphasis so far has been on the behavior of individual components and the ideal circuit elements that serve as models. Now we are ready to consider combinations of elements into circuits, using two well-known experimental laws. Our first objective is to learn how to formulate and solve circuit equations.

Circuits of considerable complexity or generality are called *networks* and circuit principles capable of general application are called *network theorems.* Some of these theorems are useful in reducing complicated networks to simple ones. Other theorems enable us to draw general conclusions about network behavior. This chapter focuses on networks consisting of resistances and steady (dc) sources; later the theorems will be extended to include networks containing other elements and other sources.

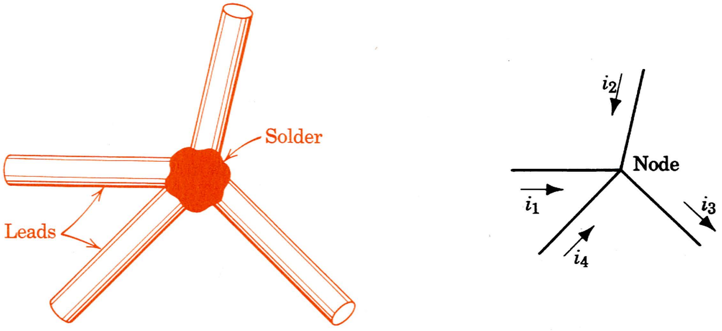
We start with *linear* networks, i.e., combinations of components that can be represented by the ideal circuit elements R, L, and C, and ideal energy sources. However, most electronic devices and many practical components are nonlinear. Therefore, we consider how linear methods can be used in analyzing some nonlinear networks, and we learn some new techniques applicable to nonlinear devices.

Circuit Laws

The foundations of network theory were laid about 150 years ago by Gustav Kirchhoff, a German university professor, whose careful experiments resulted in the laws that bear his name. In the following discussion, a *branch* is part of a circuit with two terminals to which connections can be made, a *node* is the point where two or more branches come together, and a *loop* is a closed path formed by connecting branches.

Kirchhoff's Current Law

To repeat Kirchhoff's experiments in the laboratory, we could arrange to measure the currents in a number of conductors or "leads" soldered together (Fig. 2.1a). In every case, we would find that the sum of the currents flowing into the common point (a *node*) at any instant is just equal to the sum of the currents flowing out.



(a) A soldered connection **(b) The circuit model**

Figure 2.1 Kirchhoff's current experiment.

In Fig. 2.1b, a circuit model is used to represent an actual connection. The arrows define the *reference* direction for positive current where current is defined by the motion of positive charges. The quantity "i_1" specifies the *magnitude* of the current and its algebraic *sign* with respect to the reference direction. If i_1 has a value of "+5 A," the effect is as if positive charge is moving toward the node at the rate of 5 C/s. If $i_1 = +5$ A flows in a metallic conductor in which charge is transported in the form of negative electrons, the electrons are actually moving away from the node but the effect is the same. If i_2 has a value of "−3 A," positive charge is in effect moving away from the node. In a practical situation the direction of current flow is easy to determine; if an ammeter reads "upscale," current is flowing from the meter terminal marked + through the meter to the terminal marked −.

By Kirchhoff's current law, *the algebraic sum of the currents into a node at any instant is zero.* It is sometimes convenient to abbreviate this statement, and write "$\Sigma i = 0$," where the Greek letter sigma stands for "summation." As applied to Fig. 2.1b,

$$\Sigma i = 0 = i_1 + i_2 - i_3 + i_4 \tag{2-1}$$

Obviously some of the currents may be negative.

Example 1

In Fig. 2.2, if $i_1 = +5$ A, $i_2 = -3$ A, and $i_4 = +2$ A, what is the value of i_3?

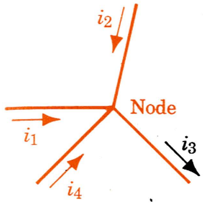


Figure 2.2 Current calculation.

By Kirchhoff's current law,

$$\Sigma i = 0 = i_1 + i_2 - i_3 + i_4$$
$$\Sigma i = 0 = +5 - 3 - i_3 + 2$$

or

$$i_3 = +5 - 3 + 2 = +4\text{ A}$$

Note that we could just as well say that the algebraic sum of the currents *leaving* a node is zero.

Kirchhoff's Voltage Law

The current law was originally formulated on the basis of experimental data. The same result can be obtained from the principle of conservation of charge and the definition of current. Kirchhoff's voltage law also was based on experiment, but the same result can be obtained from the principle of conservation of energy and the definition of voltage.

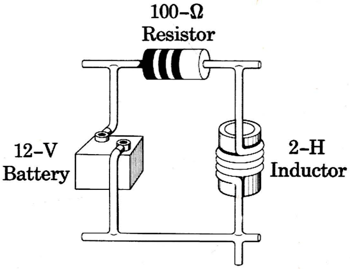


Figure 2.3 A voltage experiment.

To repeat Kirchhoff's observations about voltages, we could set up an electrical circuit and arrange to measure the voltages across a number of components that form a closed path. Only a portion of the circuit is shown in Fig. 2.3, but the combination of a battery, a resistor, an inductor, and the associated leads forms the desired closed path. In every case, we would find that the voltages around the loop, when properly combined, add up to zero.

The circuit model of Fig. 2.4 is more convenient to work with. There the voltages are labeled and the + signs define the *reference* direction for positive voltage or potential difference. The quantity v_S specifies the *magnitude* of the source voltage and its algebraic *sign* with respect to the reference. If v_S has a value of +12 V, the voltage of node b with respect to node a is positive. Since, by definition, voltage is energy per unit charge, a positive charge of 1 C moving from node a to b gains 12 J of energy from the voltage source. If v_R has a value of +5 V, a positive charge of 1 C moving from b to c loses 5 J of energy; this energy is removed from the circuit and dissipated in the resistance R. If v_L has a value of -7 V, node d is at a higher voltage than c and a positive charge of 1 C moving from c to d *gains* 7 J of energy. A current flowing from c to d could receive energy previously stored in the magnetic field of inductance L.

By Kirchhoff's voltage law, *the algebraic sum of the voltages around a loop at any instant is zero.* As an abbreviation, we may write "$\Sigma v = 0$." As applied to Fig. 2.4,

$$\Sigma v = 0 = v_{ba} + v_{cb} + v_{dc} + v_{ad} \tag{2-2}$$

where v_{ba} means "the voltage of b with respect to a"; if v_{ba} is positive, terminal b is at a higher potential than terminal a. In applying the law, the loop should be traversed in one continuous direction, starting at one point and returning to the same point. Starting from node a in this case,

$$\Sigma v = 0 = v_S - v_R - v_L + 0 \tag{2-3}$$

A minus sign is affixed to the v_R term because the plus sign on the diagram indicates that node b is nominally at a higher potential than c or $v_{cb} = -v_R$.

Example 2

In Fig. 2.4, if $v_S = +12$ V and $v_R = +5$ V, what is the value of v_L?

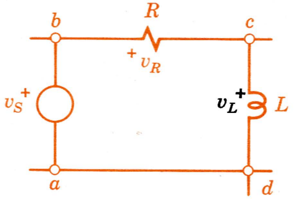


Figure 2.4 Voltage calculation.

By Kirchhoff's voltage law,

$$\Sigma v = 0 = v_S - v_R - v_L + 0$$

$$\Sigma v = 0 = +12 - 5 - v_L + 0$$

or

$$v_L = +12 - 5 = +7 \text{ V}$$

The loop can be traversed in either direction, starting at any point. Going counterclockwise, starting at node c,

$$\Sigma v = 0 = +v_R - v_S + 0 + v_L$$

which agrees with Eq. 2-3.

Application of Kirchhoff's Laws

To illustrate the application of Kirchhoff's laws in solving electric circuits, consider the circuit shown in Fig. 2.5. In this case, the source voltages and the resistances are given and the element currents and voltages are to be determined.

The first step is to label the unknown currents arbitrarily; as drawn, the arrow to the right indicates the *reference direction* of i_1. If the value of i_1 is calculated and found to be positive, current i_1 actually flows to the right and an ammeter inserted between node a and the 2-Ω resistor with the + terminal at a would read *upscale* or positive. If the value of i_1 is found to be negative, current i_1 actually flows to the left (the ammeter inserted as previously would read *downscale* or negative).

Applying Kirchhoff's current law to node a,

$$\Sigma i_a = 0 = +i_4 - i_1$$

or

$$i_4 = i_1$$

From this we conclude that the current is the same at every point in a series circuit. Since $i_4 \equiv i_1$, it is not another unknown and will be considered no further. Note that i_2 is the current in the 20-V source as well as in the 4-Ω resistance.

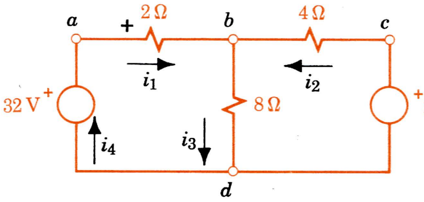


Figure 2.5 Application of Kirchhoff's laws.

The next step is to define the unknown element voltages in terms of the arbitrarily assumed currents. In flowing from a to b, the positive charges constituting current i_1 lose energy in the 2-Ω resistance; this loss in energy indicates that the potential of a is higher than that of b, or $v_{ab} = Ri_{ab} = +2i_1$. The left-hand terminal of the 2-Ω resistance is then marked with a neat + to indicate the polarity of the element voltage *in terms of the assumed current direction.* Following the same analysis, the upper end of the 8-Ω resistance and the right-hand end of the 4-Ω resistance could be marked + in accordance with the following *element equations:*

$$\begin{aligned} v_{ab} &= +2i_1 \quad &\text{or} \quad v_{ba} &= -2i_1 \\ v_{bd} &= +8i_3 \quad &\text{or} \quad v_{db} &= -8i_3 \\ v_{cb} &= +4i_2 \quad &\text{or} \quad v_{bc} &= -4i_2 \end{aligned} \tag{2-4}$$

Additional relations are obtained in the form of *connection equations.* Applying Kirchhoff's current law to node b,

$$\Sigma i_b = 0 = +i_1 + i_2 - i_3 \tag{2-5}$$

Since we are summing the currents *into* node b, i_3 is shown with a minus sign. Applying the current law to node d,

$$\Sigma i_d = 0 = -i_1 - i_2 + i_3$$

Note that this equation is *not independent;* it contributes no new information.

Applying the voltage law to the left-hand loop *abda,* and considering each element in turn,

$$\Sigma v = 0 = v_{ba} + v_{db} + v_{ad} \tag{2-6}$$

For the right-hand loop *bcdb,*

$$\Sigma v = 0 = v_{cb} + v_{dc} + v_{bd} \tag{2-7}$$

Kirchhoff's voltage law applies to any closed path; for the outside loop *abcda*,

$$\Sigma v = 0 = v_{ba} + v_{cb} + v_{dc} + v_{ad}$$

But this equation is just the sum of Eqs. 2-6 and 2-7 and no new information is obtained. Although this equation is not independent, it can be valuable in checking.

It is always possible to write as many independent equations as there are unknowns. For a circuit with six unknowns (three voltages and three currents), we have written six equations (three element and three connection equations). Other equations could be written but they would contribute no new information. If the currents are of primary interest, the unknown voltages may be eliminated by substituting Eqs. 2-4 into Eqs. 2-6 and 2-7, which become

$$\begin{aligned} \Sigma v_{abda} &= 0 = -2i_1 - 8i_3 + 32 \\ \Sigma v_{bcdb} &= 0 = +4i_2 - 20 + 8i_3 \end{aligned}$$

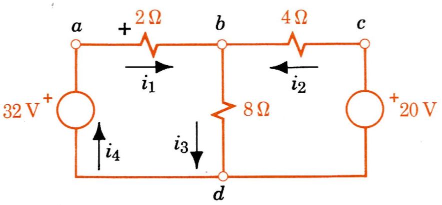


Figure 2.6 Writing circuit equations.

When these are rewritten along with Eq. 2-5, we have for Fig. 2.6

$$i_1 + i_2 - i_3 = 0 \qquad (2\text{-}8)$$

$$+2i_1 \qquad + 8i_3 = 32 \qquad (2\text{-}9)$$

$$+ 4i_2 + 8i_3 = 20 \qquad (2\text{-}10)$$

To evaluate the unknowns, these three equations are to be solved simultaneously. Some commonly employed methods are illustrated in the solution of this problem.

Solution by Determinants

The method of determinants is valuable because it is systematic and general; it can be used to solve complicated problems or to prove general theorems. The first step is to write the equations in the *standard form* of Eqs. 2-8, 2-9, and 2-10 with the constant terms on the right and the corresponding current terms aligned on the left.

A determinant is an array of numbers or symbols; the array of the coefficients of the current terms is a *third-order* determinant

$$D = \begin{vmatrix} 1 & 1 & -1 \\ 2 & 0 & 8 \\ 0 & 4 & 8 \end{vmatrix} \qquad (2\text{-}11)$$

The value of a determinant of second order is

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1 \qquad (2\text{-}12)$$

The value of a determinant of third order is

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = +a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1 \tag{2-13}$$

A simple rule for second- and third-order determinants is to take the products “from upper left to lower right” and subtract the products “from upper right to lower left.” Determinants of higher order can be evaluated by the process called “expansion by minors.”[†]

Repeating the first two columns for clarity and using the rule of Eq. 2-13,

$$D = \left|\begin{array}{ccc|cc} 1 & 1 & -1 & 1 & 1 \\ 2 & 0 & 8 & 2 & 0 \\ 0 & 4 & 8 & 0 & 4 \end{array}\right| \begin{aligned} &= +(1)(0)(8) + (1)(8)(0) + (-1)(2)(4) \\ &\quad -(1)(8)(4) - (1)(2)(8) - (-1)(0)(0) \\ &= 0 + 0 - 8 - 32 - 16 - 0 = -56 \end{aligned}$$

Replacing the coefficients of i_3 with the constant terms on the right-hand side of the equations in the standard form yields a new determinant:

$$D_3 = \begin{vmatrix} 1 & 1 & 0 \\ 2 & 0 & 32 \\ 0 & 4 & 20 \end{vmatrix} = 0 + 0 + 0 - 128 - 40 - 0 = -168 \tag{2-14}$$

By Cramer’s rule for the solution of equations by determinants,

$$i_n = \frac{D_n}{D} \quad \text{or} \quad i_3 = \frac{D_3}{D} = \frac{-168}{-56} = 3\text{ A}$$

Currents i_1 and i_2 could be obtained in a similar manner.

Substitution Method

When the number of unknowns is not large, the method of *substitution* is sometimes more convenient than using determinants. In this method, one of the equations is solved for one of the unknowns and the result is substituted in the other equations, thus eliminating one unknown. Solving Eq. 2-8 for i_3 yields

$$i_3 = i_1 + i_2$$

Substituting in Eqs. 2-9 and 2-10 yields

$$\begin{aligned} 2i_1 + 8(i_1 + i_2) &= 32 \\ 4i_2 + 8(i_1 + i_2) &= 20 \end{aligned}$$

[†] See Britton and Snively: *College Algebra,* Holt, Rinehart & Winston, New York, 1953, or any other college algebra text. An abbreviated discussion of determinants is given in the Appendix of H. H. Skilling: *Electrical Engineering Circuits,* 2nd ed., John Wiley & Sons, New York, 1965.

which can be rewritten as

$$(2+8)i_1 + 8i_2 = 32 \qquad (2\text{-}15)$$
$$8i_1 + (8+4)i_2 = 20 \qquad (2\text{-}16)$$

Solving Eq. 2-15 for i_1 yields

$$i_1 = \frac{32 - 8i_2}{10}$$

Substituting in Eq. 2-16 yields

$$8\frac{32-8i_2}{10} + 12i_2 = 20 \quad \text{or} \quad i_2 = \frac{-56}{56} = -1\text{ A}$$

and the voltage across the 4-Ω resistance is

$$v_{cb} = 4i_2 = 4(-1) = -4\text{ V}$$

The interpretation of the minus sign is that in reality node *b* is at a higher potential than *c*, and positive current flows to the right in the 4-Ω resistance.

Loop-Current Method

The reduction in the number of unknowns and in the number of simultaneous equations achieved by substitution can be obtained automatically by an ingenious approach to circuit analysis. A *loop current* I_1 is assumed to circulate around loop *abda* in Fig. 2.7, and another loop current I_2 is assumed to circulate around loop *cbdc*. By inspection, the branch current i_1 is equal to loop current I_1, and i_2 is equal to I_2; but branch current i_3 is the sum of loop currents I_1 and I_2.

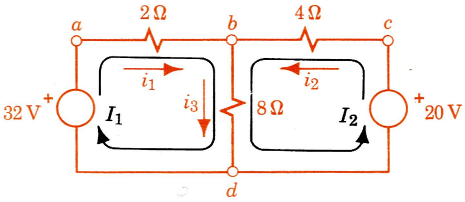


Figure 2.7 Loop currents.

By applying Kirchhoff's voltage law to the two loops we obtain

$$\Sigma v_{abda} = 0 = -2I_1 - 8(I_1 + I_2) + 32$$
$$\Sigma v_{cbdc} = 0 = -4I_2 - 8(I_1 + I_2) + 20$$

which can be rewritten as

$$(2+8)I_1 + 8I_2 = 32 \qquad (2\text{-}17)$$
$$8I_1 + (8+4)I_2 = 20 \qquad (2\text{-}18)$$

These are similar to Eqs. 2-15 and 2-16, but the physical interpretation is different. Equation 2-17 says: "The source voltage in loop 1 is equal to the sum of two voltage drops. The first, $(2+8)I_1$, is the product of loop current I_1 and the sum of all the resistances in loop 1. The second, $8I_2$, is the product of loop current I_2 and the sum of all resistances that are common to loops 1 and 2." The *self-resistance* of loop 1 is $2 + 8 = 10\,\Omega$; the *mutual resistance* between the two loops is $8\,\Omega$. A similar statement could be made by using Eq. 2-18. To obtain I_1, multiply Eq. 2-17 by 3 and Eq. 2-18 by (-2) and add the resulting equations. Then

$$\begin{array}{rcr} 30I_1 + 24I_2 & = & 96 \\ -16I_1 - 24I_2 & = & -40 \\ \hline 14I_1 \qquad\quad & = & 56 \end{array} \quad \text{or} \quad I_1 = 4\text{ A} = i_1$$

Checking

In solving complicated circuits there are many opportunities for mistakes and a reliable check is essential. This is particularly true when the solution is obtained in a mechanical way, using determinants. A new equation such as that obtained by writing Kirchhoff's voltage law around the outside loop of Fig. 2.7 provides a good check. With a little experience you will be able to write (by inspection, without first writing the element equations)

$$\Sigma v = 0 = -2i_1 + 4i_2 - 20 + 32$$

Substituting $i_1 = 4$ A and $i_2 = -1$ A,

$$\Sigma v = 0 = -2(4) + 4(-1) - 20 + 32 = 0$$

and the values of i_1 and i_2 are checked.

Example 3

Given the circuit of Fig. 2.8, calculate the current in the 10-Ω resistance using loop currents.

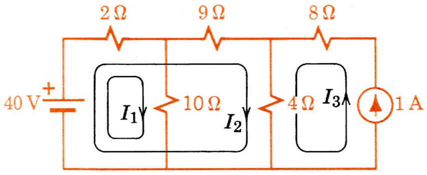


Figure 2.8 Loop-current analysis.

1. Assume loop currents so that the desired current is I_1 by definition and $I_3 = 1$ A as given.
2. Applying Kirchhoff's voltage law to two loops,

$$\begin{aligned} \Sigma v &= 0 = 40 - 2(I_1 + I_2) - 10I_1 \\ \Sigma v &= 0 = 40 - 2(I_1 + I_2) - 9I_2 - 4(I_2 + 1) \end{aligned}$$

In terms of self and mutual resistance,

$$(10 + 2)I_1 + 2I_2 = 40 \qquad (2\text{-}19)$$

$$2I_1 + (2 + 9 + 4)I_2 + 4 \times 1 = 40 \qquad (2\text{-}20)$$

From Eq. 2-20, $I_2 = (36 - 2I_1)/15$.

Substituting into Eq. 2-19,

$$12I_1 + (72 - 4I_1)/15 = 40$$

or

$$I_1 = \frac{600 - 72}{180 - 4} = \frac{528}{176} = 3\text{ A}$$

Node-Voltage Method

The wise selection of loop currents can significantly reduce the number of simultaneous equations to be solved in a given problem. In the preceding example, the number of equations was reduced from three to two by using loop currents. Would it be possible to get all the important information in a single equation with a single unknown? In this particular case, the answer is "yes" if we choose as the unknown the voltage of node b with respect to a properly chosen reference. In some practical devices many components are connected to a metal "chassis," which in turn is "grounded" to the earth. Such a ground, often shown as a common lead at the bottom of a circuit diagram, is a convenient reference. In this problem the greatest simplification will result if we choose node d; now the voltage of any node is understood to be with respect to node d, i.e., the voltage of node d is zero. (See Fig. 2.9.)

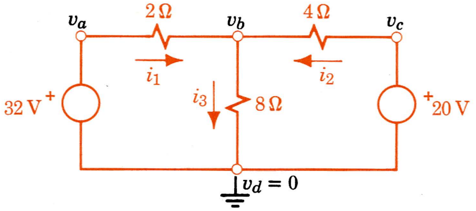


Figure 2.9 Node voltages.

Next, apply Kirchhoff's current law to each independent node. Here there is only one independent node (since the voltages of nodes a and c are fixed by the voltage sources), so the sum of the currents into node b is

$$\Sigma i_b = 0 = i_1 + i_2 - i_3 \tag{2-21}$$

The current i_1 to node b from node a is equal to the difference in potential between nodes a and b divided by the resistance between a and b, or

$$i_1 = \frac{v_a - v_b}{R_{ab}} \tag{2-22}$$

Similarly,

$$i_2 = \frac{v_c - v_b}{R_{cb}} \quad \text{and} \quad i_3 = \frac{v_b - 0}{R_{bd}} \tag{2-23}$$

where all voltages are in reference to node d. In this particular problem, v_a is given as $+32$ V, and v_c is given as $+20$ V.

Finally, bearing in mind Eq. 2-21 and the concepts represented by Eqs. 2-22 and 2-23, we write one equation with a single unknown,

$$\Sigma i_b = 0 = \frac{32 - v_b}{2} + \frac{20 - v_b}{4} - \frac{v_b - 0}{8} \tag{2-24}$$

Multiplying through by 8,

$$128 - 4v_b + 40 - 2v_b - v_b = 0$$

Therefore, the unknown voltage at independent node b is

$$v_b = \frac{168}{7} = 24\text{ V}$$

By inspection of Fig. 2.9 we can write

$$i_3 = \frac{24}{8} = 3\text{ A}, \quad i_1 = \frac{32-24}{2} = 4\text{ A}, \quad \text{and} \quad i_2 = \frac{20-24}{4} = -1\text{ A}$$

which agree with the results previously obtained. As illustrated in Example 4, the node-voltage approach can be applied to more complicated circuits.

Example 4

Given the circuit of Fig. 2.10, calculate the current in the 10-Ω resistance using node voltages.

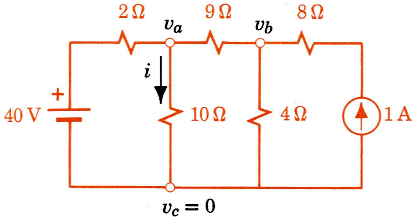


Figure 2.10 Node-voltage analysis.

1. Choose node c (one end of the branch of interest) as the reference node.
2. Apply Kirchhoff's current law to independent nodes a and b.

$$\Sigma i_a = 0 = \frac{40-v_a}{2} + \frac{v_b-v_a}{9} - \frac{v_a-0}{10}$$

$$\Sigma i_b = 0 = \frac{v_a-v_b}{9} + 1 - \frac{v_b-0}{4}$$

Multiplying through by 90 and 36, respectively,

$$\begin{aligned} 1800 - 45v_a + 10v_b - 10v_a - 9v_a &= 0 \\ 4v_a - 4v_b + 36 - 9v_b &= 0 \end{aligned}$$

Collecting terms,

$$\begin{aligned} 64v_a - 10v_b &= 1800 \\ 4v_a - 13v_b &= -36 \end{aligned}$$

3. Solving by determinants,

$$v_a = \frac{D_a}{D} = \frac{\begin{vmatrix} 1800 & -10 \\ -36 & -13 \end{vmatrix}}{\begin{vmatrix} 64 & -10 \\ 4 & -13 \end{vmatrix}} = \frac{-23{,}400 - 360}{-832 + 40}$$

$$\therefore v_a = 30\text{ V} \quad \text{and} \quad i_{10} = v_a/10 = 3\text{ A}$$

Formulation of Equations

Many books are devoted to electrical network analysis at intermediate and advanced levels; these books give a thorough treatment of a variety of methods for solving networks of great complexity and generality. Although many significant problems can be solved by using the three approaches outlined in this

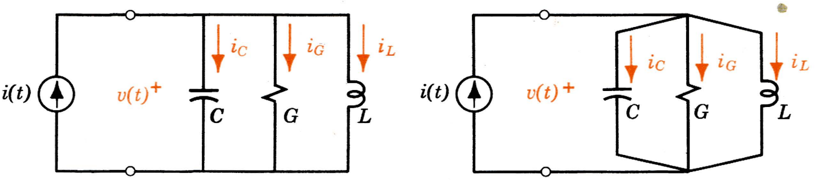


Figure 2.11 A parallel CGL circuit drawn in two ways.

chapter, the discussion here is intended to be introductory and illustrative rather than comprehensive.

At this point you should be familiar with electrical quantities (their definitions, units, and symbols), you should understand the distinction between real circuit components and idealized circuit elements, you should know the voltage-current characteristics of active and passive circuit elements, and you should be able to apply Kirchhoff's laws to obtain circuit equations. In engineering, the solution of equations is usually less demanding than the formulation, and it is proper to place primary emphasis on the formulation of equations.

If voltages and currents are unchanging with time, as in the preceding example, the behavior of a circuit is determined by resistance alone. A more general problem is illustrated in Fig. 2.11, which resembles the tuning circuit found in every transistor radio. Following the procedure used previously, we write element equations and then connection equations. Assuming a voltage $v(t)$ with the reference polarity indicated by the + sign, the corresponding current directions are as shown. Assuming no initial current in the inductance ($i_L = 0$ at $t = 0$), the element equations are:

$$i_C = C\frac{dv}{dt}, \quad i_G = Gv, \quad i_L = \frac{1}{L}\int_0^t v\,dt + 0$$

Applying Kirchhoff's current law to the upper node to obtain the connection equation,

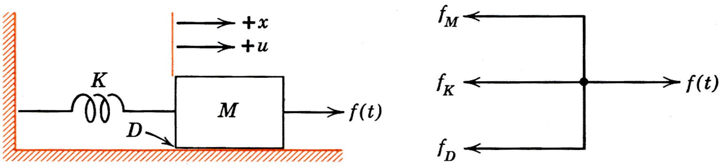
$$\Sigma i = 0 = i(t) - i_C - i_G - i_L$$

Substituting and rearranging,

$$C\frac{dv}{dt} + Gv + \frac{1}{L}\int_0^t v\,dt = i(t) \tag{2-25}$$

If the element values and $i(t)$ are known, the solution of this *integro-differential* equation consists in finding $v(t)$. In the general case, this can be quite difficult (see Chapter 6).

As another illustration, consider the dynamics problem of Fig. 2.12a. A mass restrained by a spring slides on a friction surface under the action of an applied



(a) Mechanical system (b) Freebody diagram

Figure 2.12 A dynamics problem.

force. The similarity between an electric circuit and this *mechanical circuit* is made clear by the application of d'Alembert's principle. This principle, which changes the dynamics problem into a statics problem, states that the resultant of the external forces acting on a body and the inertia force equals zero. In other words, if the inertia force (a reaction) is considered in the same manner as external forces, the sum of the forces acting equals zero, or $\Sigma f = 0$.

Again we write "element" equations and "connection" equations. For positive displacement to the right, the force f_K of a linear spring is to the left and proportional to the displacement. The friction force f_D may be assumed to be proportional to velocity and in such a direction as to oppose motion. The inertia force f_M is proportional to the mass and in such a direction as to oppose acceleration. Assuming no initial displacement and no initial velocity, the element equations are:

$$f_K = \frac{1}{K}x = \frac{1}{K}\int_0^t u\,dt + 0, \quad f_D = Du, \quad f_M = Ma = M\frac{du}{dt}$$

directed as shown in Fig. 2.12b. The connection equation, obtained by applying d'Alembert's principle, is:

$$\Sigma f = 0 = f(t) - f_M - f_D - f_K$$

Substituting and rearranging,

$$M\frac{du}{dt} + Du + \frac{1}{K}\int_0^t u\,dt = f(t) \tag{2-26}$$

If the element values and $f(t)$ are known, the solution of this integro-differential equation consists in finding $u(t)$.

Two conclusions can be drawn from these illustrations. First, there is a basic procedure in formulating equations that is applicable to a variety of electrical circuits and to problems in other fields as well. Second, for even simple circuits the equations are formidable if the variables are functions of time. If practical problems are to be solved, it appears that special mathematical tools and techniques are necessary.

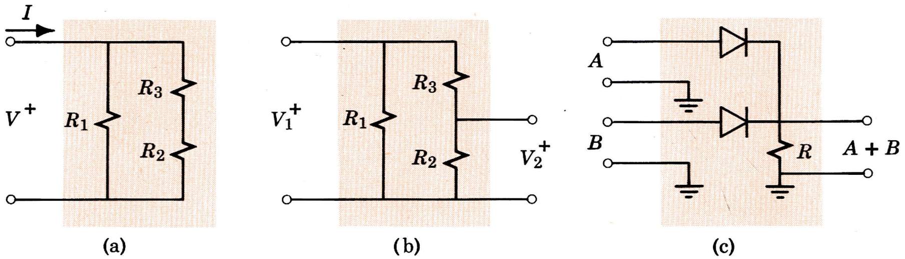


Figure 2.13 One-port, two-port, and multiport networks.

Network Theorems

A common problem in circuit analysis is to determine the response of a two-terminal circuit to an input signal. In Fig. 2.13a there is one "port" at which excitation is applied and response is measured. If the desired response is measured across R_2 as in Fig. 2.13b, this circuit becomes a *three-terminal circuit* (one terminal is common to input and output) or a *two-port circuit,* or simply a *two-port.* Transistors, transformers, and transmission lines are two-ports; typically there is an input port and a separate output port. The two-input diode **OR** gate of Fig. 2.13c is a *three-port.*

Equivalence

First we consider how complicated two-terminal circuits can be reduced to simpler circuits which are *equivalent.* By definition:

Two one-ports are equivalent if they present the same v-i characteristics.

Two one-ports consisting of various resistances connected in series and parallel are equivalent if they have the same input resistance or the same input conductance. To calculate the response I to an input V in Fig. 2.13a, what single resistance could replace the three resistances?

Network Reduction

Replacing a complicated network with a simpler equivalent is advantageous in network analysis. Applying Kirchhoff's voltage law to the circuit of Fig. 2.14a, ("closing the loop" by going through V),

$$V = V_1 + V_2 = IR_1 + IR_2 = I(R_1 + R_2) = IR_{EQ}$$

In general, for n resistances in series the equivalent resistance is

$$R_{EQ} = R_1 + R_2 + \cdots + R_n \tag{2-27}$$

Applying Kirchhoff's current law to the circuit of Fig. 2.14b,

$$I = I_1 + I_2 = VG_1 + VG_2 = V(G_1 + G_2) = VG_{EQ}$$

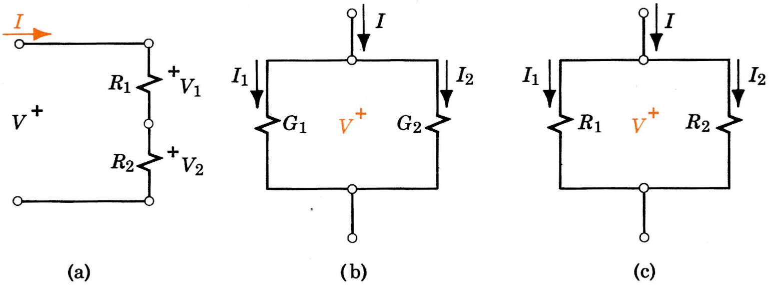


Figure 2.14 Resistances in series and parallel.

In general, for n conductances in parallel the equivalent conductance is

$$G_{EQ} = G_1 + G_2 + \cdots + G_n \tag{2-28}$$

For two resistances in parallel (Fig. 2.14c), a convenient relation is

$$R_{EQ} = \frac{1}{G_{EQ}} = \frac{1}{G_1 + G_2} = \frac{1}{(1/R_1) + (1/R_2)} = \frac{R_1 R_2}{R_1 + R_2} \tag{2-29}$$

The corresponding relationship for two conductances in series is

$$G_{EQ} = \frac{1}{(1/G_1) + (1/G_2)} = \frac{G_1 G_2}{G_1 + G_2} \tag{2-30}$$

By following a similar approach, expressions can be derived for series and parallel combinations of inductances and capacitances (see Problems 5 and 6). For n inductances in series,

$$L_{EQ} = L_1 + L_2 + \cdots + L_n \tag{2-31}$$

For n capacitances in parallel,

$$C_{EQ} = C_1 + C_2 + \cdots + C_n \tag{2-32}$$

Other useful tools in network analysis include the *voltage divider* and the *current divider*. Referring again to Figs. 2.14a and b, we see that

$$V_2 = \frac{R_2}{R_1 + R_2} V = \frac{R_2}{R_{EQ}} V \quad \text{and} \quad I_2 = \frac{G_2}{G_1 + G_2} I = \frac{G_2}{G_{EQ}} I \tag{2-33}$$

Another useful form of the current-divider equation can be obtained by using $R = 1/G$. Then

$$I_2 = \frac{G_2}{G_1 + G_2} I = \frac{1/R_2}{(1/R_1) + (1/R_2)} I \cdot \frac{R_1 R_2}{R_1 R_2} = \frac{R_1}{R_1 + R_2} I \tag{2-34}$$

When a particular voltage or current in a circuit is to be determined, using these relations is preferable to writing and solving loop or node equations.

Example 5

In the circuit of Fig. 2.15, determine the current in the 12-Ω resistance.

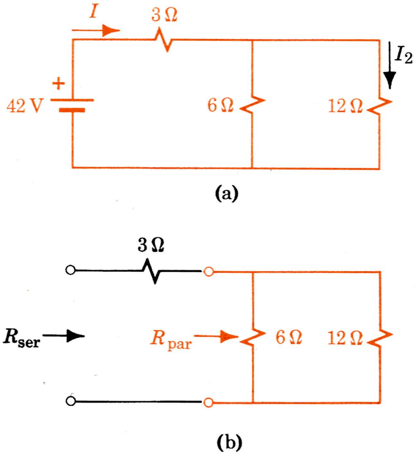


Figure 2.15 Current-divider application.

For resistances in parallel,

$$R_{par} = \frac{R_1R_2}{R_1 + R_2} = \frac{6 \times 12}{6 + 12}$$

For resistances in series,

$$R_{ser} = R + R_{par} = 3 + \frac{6 \times 12}{6 + 12}$$

Using the current-divider principle,

$$I_2 = \frac{R_1}{R_1 + R_2} I = \frac{R_1}{R_1 + R_2} \cdot \frac{V}{R_{ser}}$$

$$= \frac{6}{6 + 12} \cdot \frac{42}{3 + (6 \times 12)/(6 + 12)}$$

$$= \frac{1}{3} \cdot \frac{42}{3 + 4} = 2 \text{ A}$$

Thévenin's Theorem

The one-port networks discussed so far have been *passive;* they absorb energy from the source. In contrast, *active* networks include energy sources. A valuable method for representing active networks by simpler equivalents is described in the following statement of *Thévenin's theorem:*

> *Insofar as a load is concerned, any one-port network of resistance elements and energy sources can be replaced by a series combination of an ideal voltage source V_T and a resistance R_T where V_T is the open-circuit voltage of the one-port and R_T is the ratio of the open-circuit voltage to the short-circuit current.*

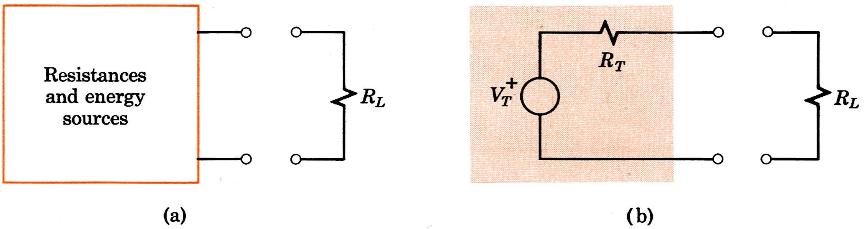


Figure 2.16 A one-port network and the Thévenin equivalent.

If the two networks of Fig. 2.16 are to be equivalent for all values of load resistance, they must be equivalent for extreme values such as $R_L = \infty$ and $R_L = 0$. The value $R_L = \infty$ corresponds to the open-circuit condition; by comparison of the two networks, the open-circuit voltage V_{OC} of the original network is equal to V_T of the equivalent. The value $R_L = 0$ corresponds to the short-circuit condition; by comparison of the two networks, the short-circuit current I_{SC} of the original network is equal to V_T/R_T of the equivalent. Therefore, the equations

$$V_T = V_{OC} \quad \text{and} \quad R_T = \frac{V_T}{I_{SC}} = \frac{V_{OC}}{I_{SC}} \tag{2-35}$$

define the components of the Thévenin equivalent network.

Alternatively, R_T is the "resistance seen by looking in" at the terminals with all independent† energy sources removed. Voltage sources must be removed by short-circuiting, and current sources must be removed by open-circuiting. By way of explanation, the resistance "seen" by the network in Fig. 2.16b is R_L; the resistance "looking in" at the terminals of the Thévenin equivalent when V_T is short-circuited is R_T.

This theorem is particularly useful when the load is to take on a series of values or when a general analysis is being performed with literal numbers. We shall use it in analyzing such devices as transistor biasing circuits and feedback amplifiers. No general proof of the theorem is given here, but Example 6 on p. 46 is a demonstration of its validity. (As a check on the result, solve for I by the loop-current method.)

The significance of the restriction "insofar as a load is concerned" is revealed if we attempt to use the Thévenin equivalent to determine internal behavior in Example 6. The power developed internally in the Thévenin equivalent is $P_T = V_T I = 21 \times 3 = 63$ W. In the original circuit, $I_1 = \frac{4}{3}$ A and the total power supplied is $P = 35 \times 3 + 42 \times \frac{4}{3} = 161$ W. The Thévenin circuit is "equivalent" only in a restricted sense.

Norton's Theorem

Our experience with networks leads us to expect that there is a corollary to Thévenin's theorem. An alternative proposition is described in the following statement called *Norton's theorem:*

> *Insofar as a load is concerned, any one-port network of resistance elements and energy sources can be replaced by a parallel combination of an ideal current source I_N and a conductance G_N where I_N is the short-circuit current of the one-port and G_N is the ratio of the short-circuit current to the open-circuit voltage.*

---

† All energy sources considered up to this point have been *independent* of voltages and currents in other parts of the circuit. *Controlled* sources, for example those used in representing transistors (Chapter 17), must not be removed in applying Thévenin's theorem.

Example 6

Using Thévenin's theorem, determine the current in the 3-Ω resistance of Fig. 2.17a.

The load resistance of interest is isolated as a load (Fig. 2.17b), and the Thévenin equivalent is drawn (Fig. 2.17c). Once the components of the equivalent circuit are determined, I is easily calculated. Under open-circuit conditions (Fig. 2.17d), the voltage divider concept indicates that the voltage across the 6-Ω resistance is

$$V_6 = \frac{6}{6+12}42 = 14\text{ V}$$

and, by Eq. 2-35,

$$V_T = V_{OC} = +14 - 35 = -21\text{ V}$$

Under short-circuit conditions (Fig. 2.17e), the loop-current equations are

$$\begin{aligned} 18I_1 - 6I_2 &= 42 \\ -6I_1 + 6I_2 &= -35 \end{aligned}$$

Solving,

$$I_2 = I_{SC} = -21/4\text{ A}$$

Hence, by Eq. 2-35,

$$R_T = \frac{V_{OC}}{I_{SC}} = \frac{-21}{-21/4} = 4\,\Omega$$

Alternatively, the resistance looking into Fig. 2.17f is

$$R_T = R_{in} = \frac{6 \times 12}{6+12} = 4\,\Omega$$

Therefore (Fig. 2.17c),

$$I = \frac{V_T}{R_T + 3} = \frac{-21}{4+3} = -3\text{ A}$$

and the current in the 3-Ω resistance is 3 A upward.

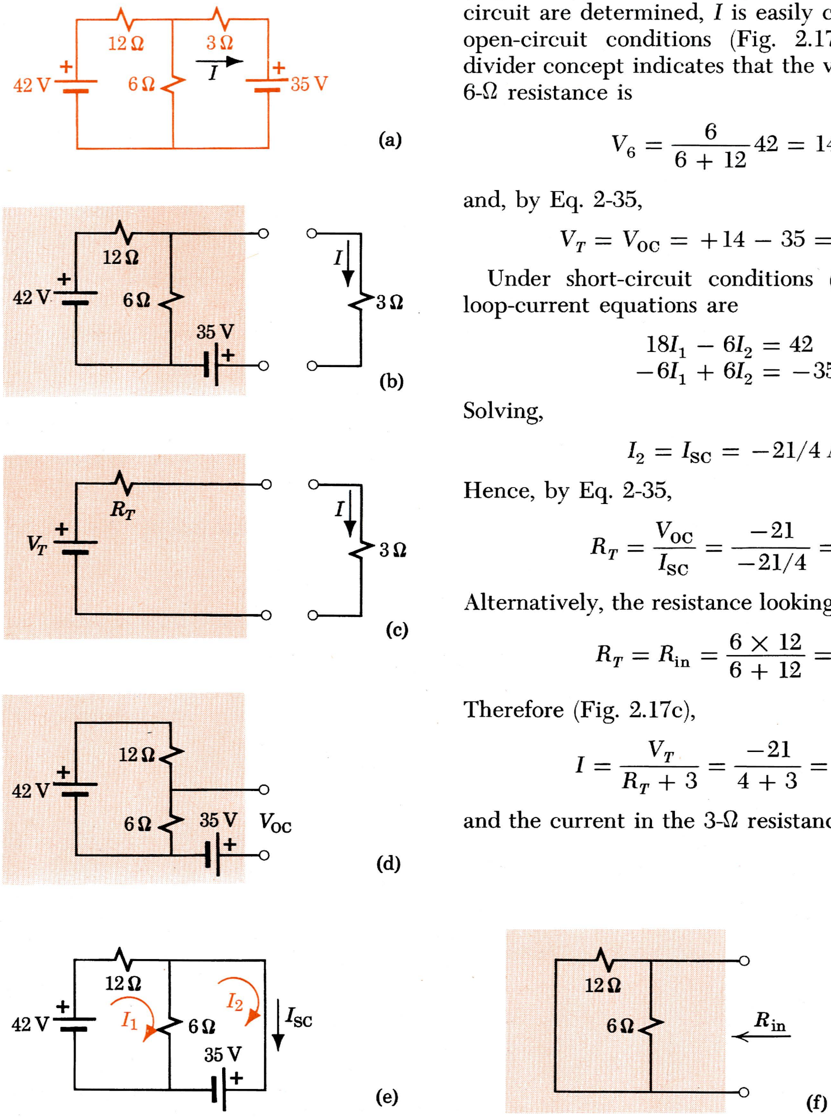


Figure 2.17 Application of Thévenin's theorem.

If the two networks of Fig. 2.18 are to be equivalent for all values of load conductance, they must be equivalent for extreme values such as $G_L = \infty$ and $G_L = 0$. The value $G_L = \infty$ corresponds to the short-circuit condition; by comparison of the two networks, the short-circuit current I_{SC} of the original is equal to I_N of the equivalent. The value $G_L = 0$ corresponds to the open-circuit condition; by comparison of the two networks, the open-circuit voltage V_{OC} of the original network is equal to I_N/G_N of the equivalent. Therefore,

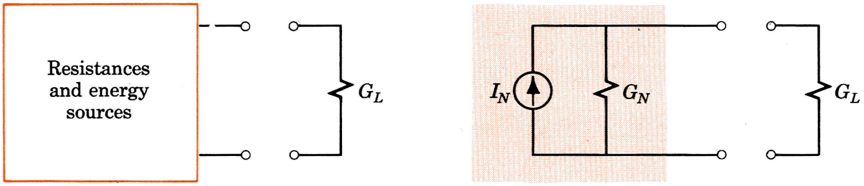


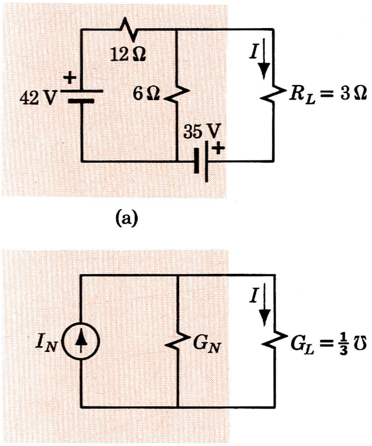
Figure 2.18 A one-port network and the Norton equivalent.

$$I_N = I_{SC} \quad \text{and} \quad G_N = \frac{I_N}{V_{OC}} = \frac{I_{SC}}{V_{OC}} \tag{2-36}$$

define the components of the Norton equivalent network. Note: $G_N = 1/R_T$.

Example 7

Using Norton's theorem, determine the current in the 3-Ω resistance of Example 6.



(a)

(b)

Figure 2.19 Application of Norton's theorem.

The circuit is redrawn in Fig. 2.19a to isolate the load.

$$V_{OC} = -21\text{ V} \quad \text{and} \quad I_{SC} = -21/4\text{ A}$$

as before. Hence, by Eqs. 2-36,

$$I_N = I_{SC} = -\frac{21}{4}\text{ A}$$

and

$$G_N = \frac{I_{SC}}{V_{OC}} = \frac{-21/4}{-21} = \frac{1}{4}\ \mho$$

By the current-divider principle,

$$I = \frac{G_L}{G_L + G_N} I_N = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{4}}\left(-\frac{21}{4}\right) = -3\text{ A}$$

as before.

As a by-product of the development of Thévenin's and Norton's theorems, we have obtained another general principle:

> *Any voltage source V with its associated series resistance R can be replaced by an equivalent current source I with an associated parallel conductance G where $I = V/R$ and $G = 1/R$.*

This principle of "source transformation" can be very useful in network analysis; for example, it permits the analyst to choose either an ideal voltage source or an ideal current source in representing a real energy source.

Superposition Theorem

The concept of *linearity* is illustrated in Fig. 2.20; an element of force ΔF produces the same element of deflection Δy at any point in the linear region.

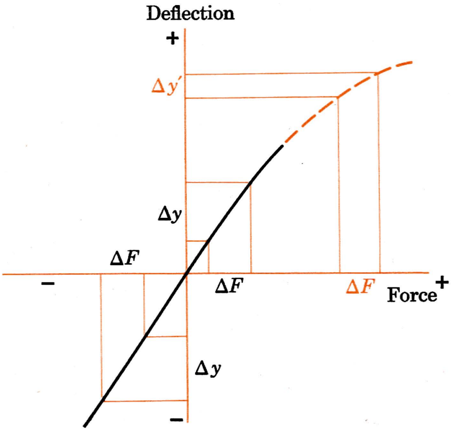


Figure 2.20 The concept of linearity.

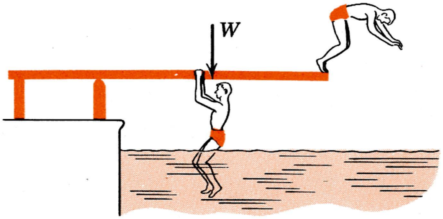


Figure 2.21 A linear beam.

In general, where effect y is a function f of cause F, or $y = f(F)$, function f is *linear* if

$$f(F + \Delta F) = f(F) + f(\Delta F) = y + \Delta y \tag{2-37}$$

In the nonlinear region, however, an equal element of force ΔF produces a different element of deflection $\Delta y'$.

In any situation where effect is directly proportional to cause, it is permissible to consider several causes individually and then combine the resulting individual effects to find the total effect. It is common practice to calculate the deflection of beams in this way. If a cantilever beam, such as the springboard of Fig. 2.21, carries three loads, the total deflection is the sum of the separate deflections due to the weight of the board itself, the swimmer hanging below, and the diver pushing on the tip. The general principle is called the *superposition theorem* and may be stated as follows:

> *If cause and effect are linearly related, the total effect of several causes acting simultaneously is equal to the sum of the effects of the individual causes acting one at a time.*

In electrical circuits, the causes are excitation voltages and currents and the effects are response voltages and currents. A typical case is the circuit of Fig 2.22a. If the base of the current source is chosen as the reference node, the potential of node a is I_2R_2 and the sum of the currents into a is

$$\frac{V - I_2R_2}{R_1} + I - I_2 = 0$$

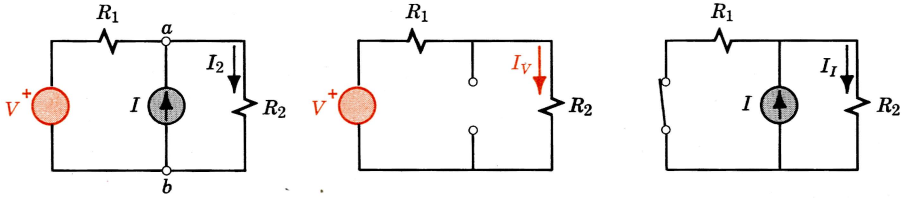
Then

$$V - I_2R_2 + IR_1 - I_2R_1 = 0$$

or

$$I_2 = \frac{V + IR_1}{R_1 + R_2} = \frac{V}{R_1 + R_2} + \frac{R_1}{R_1 + R_2}I \tag{2-38}$$

Equation 2-38 indicates that the current I_2 consists of two parts, one due to V and the other due to I. If the current source is removed (by open-circuiting) as in Fig. 2.22b, a component of current $I_V = V/(R_1 + R_2)$ flows in R_2. If the voltage source is removed (by short-circuiting) as in Fig. 2.22c, a component of current $I_I = IR_1/(R_1 + R_2)$ flows in R_2. As indicated by the superposition theorem, $I_2 = I_V + I_I$, the effects produced by the two causes separately. Note that ideal voltage sources are removed by short-circuiting and ideal current sources by open-circuiting. Real energy sources always have internal resistances that must remain in the circuit. (See Example 8 on p. 50.)



(a) Both sources (b) Voltage source (c) Current source

Figure 2.22 Superposition components.

Example 8

A load of 6 Ω is fed by a parallel combination of two batteries. Battery A has an open-circuit voltage of 42 V and an internal resistance of 12 Ω; battery B has an open-circuit voltage of 35 V and an internal resistance of 3 Ω. Determine the current I supplied by battery B, and calculate the power dissipated internally by the battery.

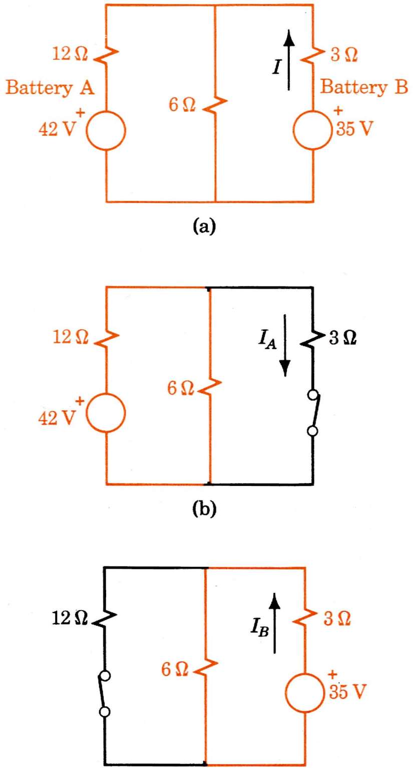


Figure 2.23 Application of the superposition principle.

If the batteries are represented by linear models, the circuit (Fig. 2.23a) is that of Examples 6 and 7. With the 35-V source removed (Fig. 2.23b), the component of current due to battery A can be written in one step as

$$I_A = \frac{6}{6+3}\,\frac{42}{\left(12 + \dfrac{3 \times 6}{3+6}\right)} = \frac{2}{3}\left(\frac{42}{12+2}\right)$$
$$= 2 \text{ A downward}$$

if the principles of current division (Eq. 2-34) and resistance combination (Eqs. 2-27 and 2-29) are applied simultaneously.

If the 42-V source is removed (the internal resistance of the battery remains), the component of current due to battery B (Fig. 2.23c) is

$$I_B = \frac{35}{3 + \dfrac{6 \times 12}{6+12}} = \frac{35}{3+4} = 5 \text{ A upward}$$

Then, by superposition,

$$I = -I_A + I_B = -2 + 5 = 3 \text{ A upward}$$

as before.

To calculate the power dissipated in the 3-Ω resistance, we might note that with the 35-V source removed, the power is

$$P_A = I_A{}^2R = 2^2 \times 3 = 12 \text{ W}$$

With the 42-V source removed, the power is

$$P_B = I_B{}^2R = 5^2 \times 3 = 75 \text{ W}$$

But the actual power dissipated is *not* 12 + 75 = 87 W. Why not? What principle would be violated in such a calculation?

The answer is that power is not linearly related to voltage or current and therefore superposition cannot be applied in power calculations. Superposition is applicable only to linear effects such as the current response. Having found the current by superposition, we can calculate the power in the 3-Ω resistance as

$$P = I^2R = (-I_A + I_B)^2R = 3^2 \times 3 = 27 \text{ W}$$

Nonlinear Networks

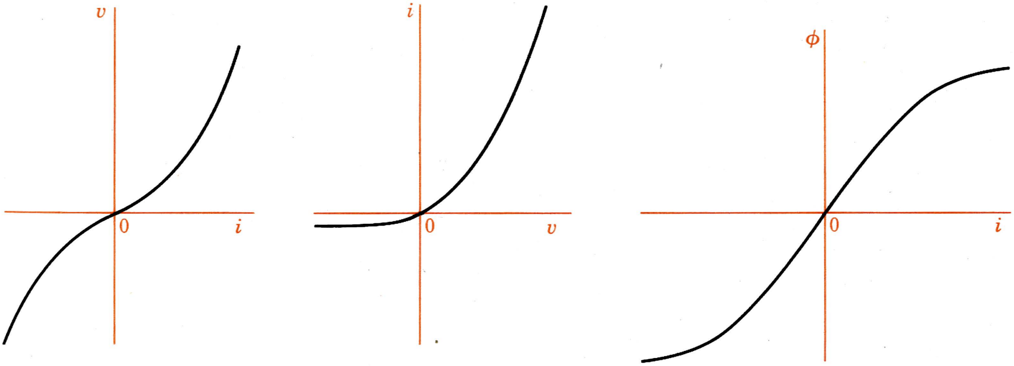
In the preceding discussions we assumed linearity, but real devices are never strictly linear. In some cases, nonlinearity can be disregarded; linear approximation yields results that predict the behavior of the real devices within acceptable limits. In other cases, nonlinearity is annoying and special steps must be taken to avoid or eliminate its effect; later we shall learn how to use feedback to minimize the distortion introduced in nonlinear electronic amplifiers. Sometimes nonlinearity is desirable or even essential; the distortion that is annoying in an amplifier is necessary in the harmonic generator for obtaining output signals at frequencies that are multiples of the input signal.

Nonlinear Elements

A dissipative element for which voltage is not proportional to current is a *nonlinear resistor.* An ordinary incandescent lamp has a characteristic similar to that in Fig. 2.24a; the "resistance" of the filament increases with temperature and, therefore, under steady-state conditions, with current.

As we shall see, most electronic devices are inherently nonlinear. Because of its v-i characteristic (Fig. 2.24b), the semiconductor *diode* is useful in discriminating between positive and negative voltages. If a sinusoidal voltage is applied to such a diode, the resulting current has a large dc component and the diode is functioning as a *rectifier.*

To increase the energy-storage capability of an inductor, an iron core with favorable magnetic properties is used. (These properties are discussed in Chapter 22.) In such an inductor, the magnetic flux ϕ is not proportional to current (Fig. 2.24c); at large values of current, a given increment of current produces only a small increment of flux. One interpretation is that "inductance" is not constant. This nonlinearity is troublesome in a power transformer, but it may be useful in a control system.



(a) Incandescent lamp (b) Semiconductor diode (c) Iron-core inductor

Figure 2.24 Characteristics of nonlinear elements.

Methods of Analysis

The method employed depends on the nature of the problem, the form of the data, and the computation aids available. The modern digital computer and analog computer permit the ready solution of problems that were hopelessly time consuming a few years ago, but here we are concerned only with simple series and parallel circuits.

Analytical Solution. If an analytical expression for the v-i characteristic can be obtained from physical principles or from experimental data, some simple problems can be solved algebraically. A useful model is the power series

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots \tag{2-39}$$

The first two terms provide a linear approximation. The first three terms give satisfactory results in many practical nonlinear problems, although more terms may be necessary in some cases.

Example 9

A voltage $v = v_m \cos \omega t$ is applied to the semiconductor diode of Fig. 2.25. Determine the nature of the resulting current.

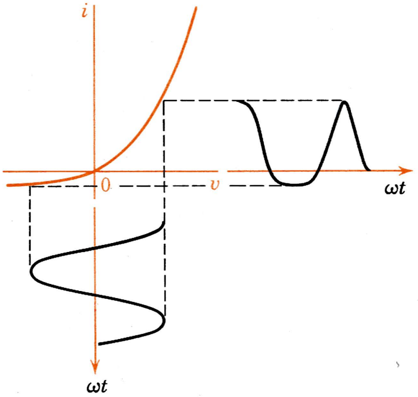


Figure 2.25 Power-series analysis.

Assuming that the v-i characteristic can be represented by the first three terms of the power series of Eq. 2-39, the current is

$$i = a_0 + a_1 v + a_2 v^2$$

But we note that for $v = 0$, $i = 0$; therefore, $a_0 = 0$. For the given voltage,

$$i = a_1 V_m \cos \omega t + a_2 V_m{}^2 \cos^2 \omega t$$

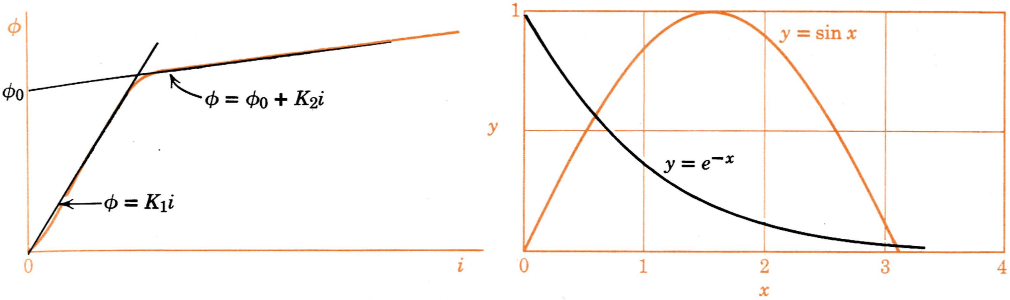
or

$$i = a_1 V_m \cos \omega t + \frac{a_2 V_m{}^2}{2}(1 + \cos 2\omega t)$$

$$= \frac{a_2 V_m{}^2}{2} + a_1 V_m \cos \omega t + \frac{a_2 V_m{}^2}{2} \cos 2\omega t \tag{2-40}$$

Equation 2-40 indicates that the resulting current consists of a steady or dc component, a "fundamental" component of the same frequency as the applied voltage, and a "second-harmonic" component at frequency 2ω.

Depending on the relative magnitudes of a_1 and a_2 and the associated circuitry, the diode of Example 9 could be used as a rectifier, an amplifier, or a harmonic generator. In a specific problem, the coefficients in the power series can be determined by choosing a number of points on the v-i characteristic equal to the number of terms to be included, substituting the coordinates of each point in Eq. 2-39, and solving the resulting equations simultaneously.



(a) Piecewise linearization

(b) Graphical solution of equations

Figure 2.26 Analysis based on graphical data.

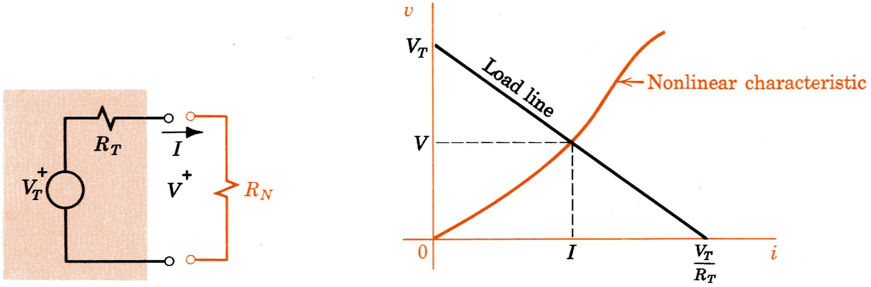
Piecewise Linearization. Where approximate results are satisfactory, a convenient approach is to represent the actual characteristic by a series of linear "pieces." The characteristic of the iron-core inductor of Fig. 2.26a can be divided into two regions, each approximated by a straight line. Below the "knee" of the curve, the behavior is adequately described by $\phi = k_1 i$; above the knee, a different expression must be used. Usually a trial solution will indicate in which region operation is occurring, and the problem is reduced to one in linear analysis. (This approach is used in developing models for electronic devices in Ch. 12.)

Graphical Solution. In piecewise linearization we use graphical data to obtain analytical expressions that hold over limited regions. Sometimes it is desirable to plot given analytical functions to obtain a solution graphically. For example, the equation $e^{-x} - \sin x = 0$ is difficult to solve analytically. If the functions $y = e^{-x}$ and $y = \sin x$ are plotted as in Fig. 2.26b, the simultaneous solutions of these two equations are represented by points whose coordinates satisfy both equations. Therefore, the intersections of the two curves are the solutions to the original equation and can be read directly from the curves. The precision of this method is limited only by the time available for plotting on successively larger scales. This method is also useful when the characteristic of some portion of the circuit is available in graphical form or in a table of experimental data.

Networks with One Nonlinear Element

If there is a single nonlinear resistance in an otherwise linear resistive network, and if the v-i characteristic of the nonlinear element is known, a relatively simple method of solution is available. This situation occurs frequently in practice and the method of attack deserves emphasis.

The first step is to replace all except the nonlinear element with the Thévenin equivalent shown in Fig. 2.27a as V_T and R_T. The combination of V_T and R_T



(a) (b)

Figure 2.27 Load-line analysis of a nonlinear network.

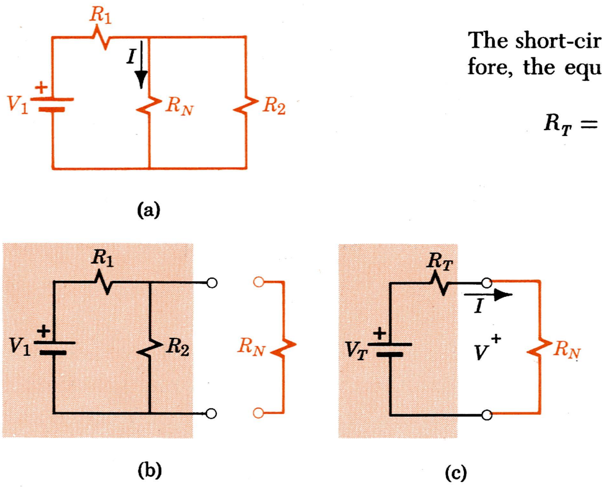
is equivalent "insofar as a load is concerned." For any value of load resistance, the terminal voltage is

$$V = V_T - R_T I \tag{2.41}$$

This equation is that of a straight line, the *load line*, with intercepts V_T and

Example 10

The nonlinear element whose characteristic is given in Fig. 2.27b is connected in the circuit of Fig. 2.28a. Determine the current I.



(a)

(b) (c)

Figure 2.28 Load-line analysis.

First, the nonlinear element is isolated (Fig. 2.28b) and the Thévenin equivalent determined. Considering R_1 and R_2 as a voltage divider,

$$V_T = V_{OC} = \frac{R_2}{R_1 + R_2} V_1$$

The short-circuit current is just $I_{SC} = V_1/R_1$. Therefore, the equivalent resistance is

$$R_T = \frac{V_{OC}}{I_{SC}} = \frac{R_2 V_1}{R_1 + R_2} \frac{R_1}{V_1} = \frac{R_1 R_2}{R_1 + R_2}$$

which is also the resistance seen looking in at the terminals with the voltage source removed.

Next, the load line is drawn with intercepts V_T and V_T/R_T as shown in Fig. 2.27b. The intersection of the load line with the nonlinear characteristic gives the desired current I.

V_T/R_T and slope $-R_T$. (For the original network, the intercepts are V_{OC} and I_{SC}.) The graph of the nonlinear characteristic is v as a function of i for the load. The simultaneous satisfaction of these two relations, which yields the values of V and I at the terminals, is the intersection of the two curves (Example 10).

For transistors, vacuum tubes, and some control system components, the *v-i* characteristics are represented by a family of curves (usually plotted i versus v). The "load" may be the linear element represented by R_T with power supplied by a battery of voltage V_T. Once the load line is drawn, the effect of variations in operating conditions (signals) is clearly visible. Note that if V_T varies during operation, the load line moves parallel to itself.

SUMMARY

◆ The algebraic sum of the currents into a node is zero ($\Sigma i = 0$).
The algebraic sum of the voltages around a loop is zero ($\Sigma v = 0$).

◆ The general procedure for formulating equations for circuits is:

1. Arbitrarily assume a consistent set of currents and voltages.
2. Write the element equations by applying the element definitions and write the connection equations by applying Kirchhoff's laws.
3. Combine the element and connection equations to obtain the governing circuit equation in terms of the unknowns.

The resulting simultaneous equations can be solved by:
Successive substitution to eliminate all but one unknown, or
Use of determinants and Cramer's rule.

◆ Use of loop currents or node voltages may greatly reduce the number of unknowns and simplify the solution.
Checking is essential because of the many opportunities for mistakes in sign and value.

◆ Two one-ports are equivalent if they present the same *v-i* characteristics.
Two passive one-ports are equivalent if they have the same input resistance.

For resistances in series, $R_{EQ} = R_1 + R_2 + \cdots + R_n$
For conductances in parallel, $G_{EQ} = G_1 + G_2 + \cdots + G_n$

◆ For the voltage divider and current divider,

$$V_2 = \frac{R_2}{R_1 + R_2} V \quad \text{and} \quad I_2 = \frac{G_2}{G_1 + G_2} I = \frac{R_1}{R_1 + R_2} I$$

◆ Insofar as a load is concerned, any one-port network of resistance elements and energy sources can be replaced by a series combination (Thévenin) of an ideal voltage source V_T and a resistance R_T, or by a parallel combination (Norton) of an ideal current source I_N and a conductance G_N, where $V_T = V_{OC}$, $R_T = V_{OC}/I_{SC}$, $I_N = I_{SC}$, and $G_N = I_{SC}/V_{OC} = 1/R_T$.

- A series combination of voltage V and resistance R can be replaced by a parallel combination of current $I = V/R$ and conductance $G = 1/R$.
- If cause and effect are linearly related, the principle of superposition applies, and the total effect of several causes acting simultaneously is equal to the sum of the effects of the individual causes acting one at a time.
- Networks involving nonlinear elements can be solved in various ways, depending on the nature of the problem and the form of the data.
  If there is a single nonlinear element in an otherwise linear network, construction of a load line permits a simple graphical solution.

REVIEW QUESTIONS

1. Write out in words a physical interpretation of Eq. 2-18.
2. Outline a three-step procedure for applying the node-voltage method.
3. Outline a three-step procedure for applying the loop-current method.
4. If two passive one-ports are equivalent, is the same power dissipated in both? Prove your answer.
5. What is the difference between "passive" and "active" networks?
6. Under what circumstances is Thévenin's theorem useful?
7. Is the same power dissipated in an active circuit and the Norton equivalent?
8. How are the Norton and Thévenin components related?
9. Cite two nonelectrical examples where superposition applies.
10. Under what circumstances is the superposition principle useful?
11. How are voltage and current sources "removed"?
12. Cite nonelectrical examples of linear and nonlinear devices.
13. Give two engineering examples where nonlinearity is desirable.
14. The v-i characteristic of an active device is $v = 100 - 2i - i^2$. Outline two methods for determining the current that flows when a resistance R is connected to the device.
15. What is meant by "piecewise linearization"? A "load line"?

EXERCISES

1. In Fig. 2.29, $i_1 = 2$ A and $i_2 = -3$ A. Predict i_3.

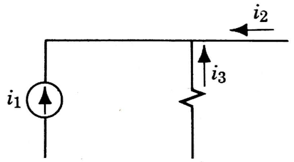


Figure 2.29

2. In Fig. 2.30, $v = 12$ V, $R_1 = 4\ \Omega$, and $R_2 = 2\ \Omega$. Calculate current i and voltages v_1 and v_2 across the resistances.
3. In Fig. 2.31, determine i_1 and i_2 in terms of the other quantities.

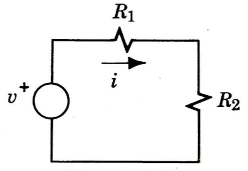


Figure 2.30

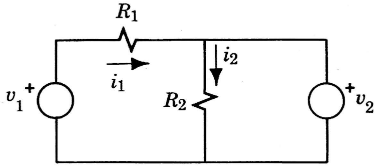


Figure 2.31

4. In Fig. 2.32, given $I_a = 3$ A, $I_b = 2$ A, and $I_c = -8$ A, determine i_1, i_2, and v.

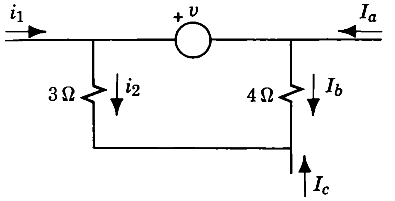

i_1
+ v
I_a
3 Ω
i_2
4 Ω
I_b
I_c

Figure 2.32

5. In Fig. 2.32, given $I_a = -4$ A, $I_c = 1$ A, and $v = 18$ V, determine i_1, i_2, and I_b.
6. Given the circuit of Fig. 2.33, calculate the voltage across the 2-Ω resistance using: (a) element currents, (b) loop currents, (c) node voltages. Include an adequate check.

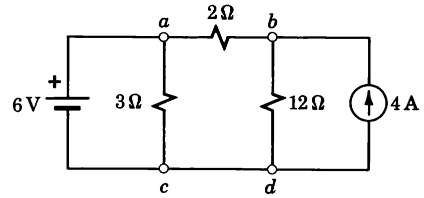

a
2 Ω
b
+
6 V
3 Ω
12 Ω
4 A
c
d

Figure 2.33

7. Given the circuit of Fig. 2.34, calculate the current in the 2-Ω resistance using: (a) element currents, (b) loop currents, (c) node voltages. Draw a conclusion about the methods.

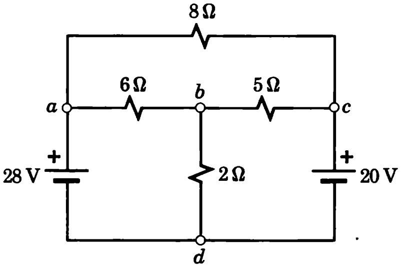

8 Ω
6 Ω
b
5 Ω
a
c
+
28 V
2 Ω
+
20 V
d

Figure 2.34

8. Given the circuit of Fig. 2.34, calculate the voltage across the 5-Ω resistance using: (a) element currents, (b) loop currents, (c) node voltages.

9. In Fig. 2.35, $i_1 = 5$ A, $i_2 = 10 \cos t$ A, and $v_C = 5 \sin t$ V. Find:
   (a) i_L and v_L
   (b) v_{ac}, v_{ab}, and v_{cd}
   (c) the energy stored in L and C, all as functions of time.

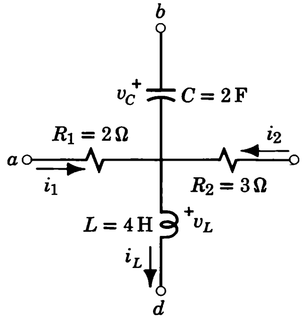

b
v_C +
$C = 2$ F
$R_1 = 2\,\Omega$
i_2
a
c
i_1
$R_2 = 3\,\Omega$
$L = 4$ H
+ v_L
i_L
d

Figure 2.35

10. Solve Exercise 9 for $i_1 = 10\,e^{-2t}$ A, $i_2 = 4$ A, and $v_C = 3\,e^{-2t}$ V.
11. In Fig. 2.36, V_0 is an initial voltage on C. Using only the specified variables, formulate the voltage equation for loop *abcda* with the switch closed.

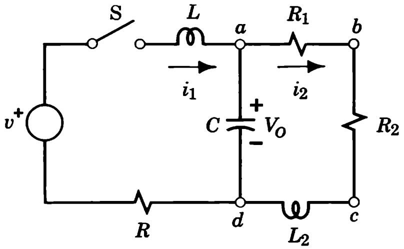

S
L
a
R_1
b
i_1
i_2
v +
+
C
V_0
−
R_2
R
d
L_2
c

Figure 2.36

12. Find current I in the *ladder* network of Fig. 2.37 by network reduction. Reduce series and parallel combinations to a single

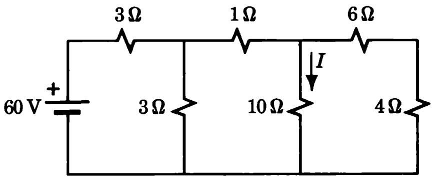

3 Ω
1 Ω
6 Ω
I
+
60 V
3 Ω
10 Ω
4 Ω

Figure 2.37

equivalent resistance, find the battery current, then use current division to find I.

13. Find current I in the *ladder* network of Fig. 2.37 by the following ingenious method: Assume the current is 1 A and work backward to find the necessary source voltage; determine actual I by simple proportion.

14. Applying the current divider principle to Fig. 2.37, write down an expression for current I as a function of V (= 60 V) and solve for I.

15. Replace the circuit of Fig. 2.37 by a simpler equivalent that will enable you to find the current in the 1-Ω resistor by a single calculation. Calculate the current.

16. In Fig. 2.38, $R_2 = 2$ kΩ, $R_1 = 10$ kΩ, and $v = 24$ V.

(a) Predict v_2.

(b) A practical voltmeter can be represented by a 10-kΩ resistance in series with an ideal (resistanceless) meter calibrated to indicate the voltage across the 10-kΩ meter resistance. Predict the reading of the voltmeter when connected across R_2.

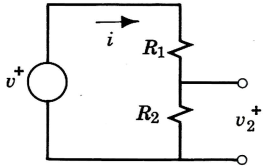


Figure 2.38 Voltage divider.

17. Repeat Exercise 16(b) if the effective voltmeter resistance is 100 kΩ, and draw a conclusion about the desirable internal resistance of a voltmeter.

18. The open-circuit voltage of an audio oscillator is 5 V. The terminal voltage drops to 4 V when a 2000-Ω resistor is connected across the terminals. Determine the internal resistance of the oscillator.

19. In Fig. 2.33, use Thévenin's theorem to find the current in the 12-Ω resistance.

20. Determine current I_1 in Fig. 2.39 using Thévenin's theorem.

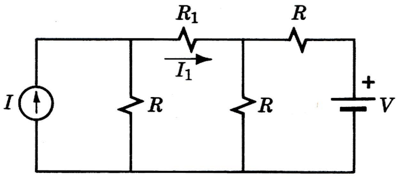


Figure 2.39

21. In Fig. 2.40, with switch S closed, the ammeter AM reads 3 A. Predict the ammeter reading with switch S open.

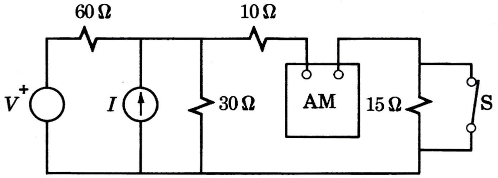


Figure 2.40

22. Repeat Exercise 6 using superposition.

23. Repeat Exercise 8 using superposition.

24. Use superposition to find current I_1 in Fig. 2.39.

25. For device A in Fig. 2.41,

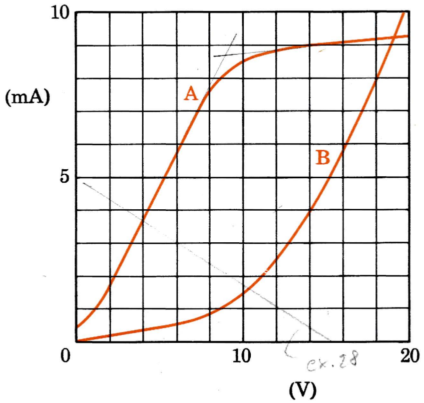


Figure 2.41

(a) Reproduce the v-i characteristic and represent it by piecewise linearization.

(b) Predict the approximate current in response to an applied voltage $v = 15 + 5\cos\omega t$ V.

26. Assume that device B in Fig. 2.41 can be represented by the equation $i(\text{mA}) = a_1 v + a_2 v^2$ in the region $10 \le v \le 18$ V.
(a) Evaluate a_1 and a_2 by simultaneous solution of two equations for i at the two values of v.
(b) Use your mathematical model to determine i for $v = 14$ V and check graphically.

27. Device A in Fig. 2.41 is connected in the circuit of Fig. 2.42 where $R = 1000\ \Omega$ and $R_1 = 600\ \Omega$.
(a) If $I = 16$ mA, predict the device current I_D.
(b) Specify the value of I so that the power dissipated in D is 60 mW.

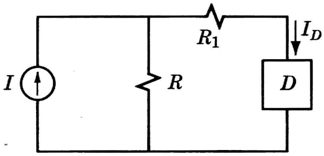


Figure 2.42

28. Device B in Fig. 2.41 is connected in parallel with a 5-mA current source and a 3-kΩ resistance; predict the voltage and the current in device B.

PROBLEMS

1. Given the circuit of Fig. 2.43, consider various methods to calculate the current in the 2-Ω resistance. Use the method that appears to require the least algebra.

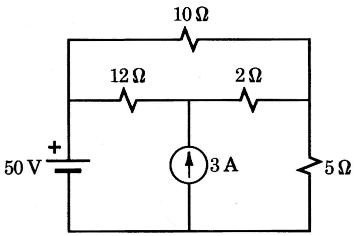


Figure 2.43

2. In Fig. 2.44, find the voltage across the 20-Ω resistor using loop currents. If possible, choose the loop currents so that only one equation is necessary.

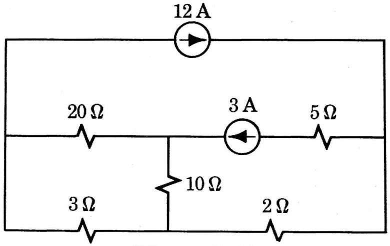


Figure 2.44

3. In the circuit of Fig. 2.45, the sources have been connected for a long time. What does this imply regarding i_C and v_L? Calculate the current in the 10-kΩ resistance, the charge on the capacitance, and the energy stored in the inductance.

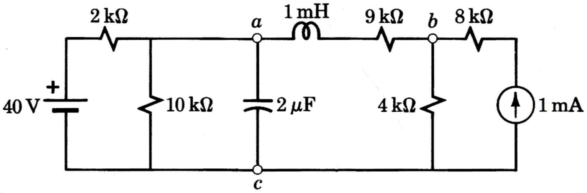


Figure 2.45

4. In Fig. 2.33, replace the portion of the circuit to the left of terminals b and d by a Norton equivalent. Then combine the Norton equivalent with the 4-A source and determine the current in the 12-Ω resistance.

5. (a) Draw and label a circuit showing inductors L_1 and L_2 connected in series and derive an equation for the equivalent inductance of the combination.
(b) Repeat for inductances L_1 and L_2 connected in parallel.

6. (a) Draw and label a circuit showing capacitors C_1 and C_2 connected in series and derive an equation for the equivalent capacitance of the combination.

(b) Repeat for capacitances C_1 and C_2 connected in parallel.

7. In Fig. 2.46, the shaded portion is one possible circuit model for a transistor. Note that 50 I_1 is a "controlled" current source. (See footnote, p. 45.) Replace the model and the voltage source by a Norton equivalent and predict the "voltage gain" V_2/V_1 for $R_L = 3\ \text{k}\Omega$.

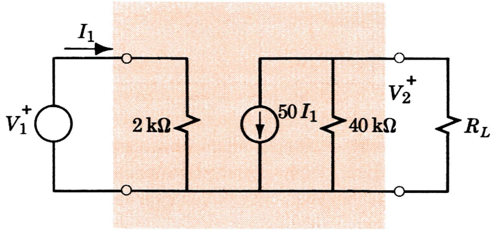


Figure 2.46

8. In an amplifier a transistor can be represented by the two-port model in Fig. 2.47, where $R_1 = 10\ \text{k}\Omega$, $R_2 = 75\ \text{k}\Omega$, $R_3 = 150\ \text{k}\Omega$, and $\beta = 100$. (βI_1 is a "controlled" current source.)

(a) Simplify the model and voltage source and predict the voltage amplification V_L/V_S as a function of R_L.

(b) Calculate V_L/V_S for $R_L = 20\ \text{k}\Omega$.

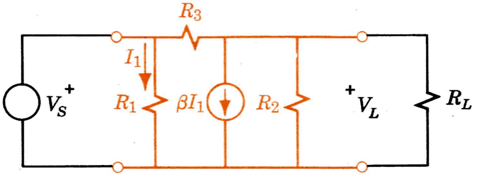


Figure 2.47

9. A signal generator, consisting of a complicated network of nearly linear passive and active elements, supplies a variable load resistance R_L.

(a) Use Thévenin's theorem to replace the generator by a combination of V_G and R_G.

(b) Prove the *maximum power transfer* theorem, which states that: "For maximum power transfer from a source to a load, the resistance of the load should be made equal to the Thévenin equivalent resistance of the source."

10. The behavior of a certain germanium diode is defined by $i = I_s(e^{bv} - 1)$ where $I_s = 10\ \mu\text{A}$ and $b = 40\ \text{V}^{-1}$.

(a) Expand the exponential term in a power series and represent $i(v)$ using the first three terms of Eq. 2.39.

(b) Predict the current if a voltage $v = +0.05$ V is applied to the diode in series with a 500-Ω resistance.

(c) Check your answer to part (b) by comparing the sum of the voltage drops across the diode and the resistance at the predicted current with the applied voltage.

11. Devices A and B in Fig. 2.41 are connected in series and the combination in series with a 30-V battery and a 5-kΩ resistance.

(a) Plot the composite v-i characteristic for A and B in series.

(b) Predict i, v_A, and v_B by the load-line method.

(c) Predict the power dissipated in devices A and B.

12. The output characteristics of two generators are:

I(A)	0	25	50	75	100
V_1(V)	120	119	117	113	105
V_2(V)	120	118	113	105	90

The two generators are connected in parallel to supply power to a load resistance $R_L = 1.0\ \Omega$. Plot the characteristics and determine graphically the current and power supplied by each generator.

Chapter 3

Signal Waveforms

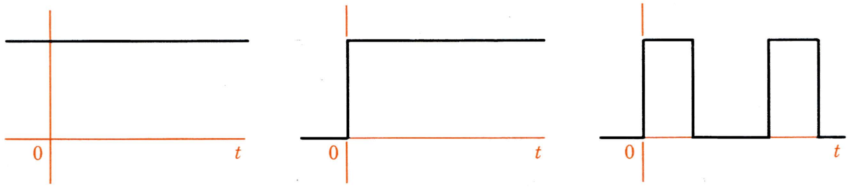
Exponentials
Sinusoids
Phasors
Periodic Waveforms

The definitions and laws of Chapters 1 and 2 provide the basis for formulating the equations governing the behavior of any circuit consisting of lumped linear elements. Well-established mathematical methods are available for solving the circuit equations, but these general methods are often quite laborious. Engineers are always interested in techniques that turn difficult problems into simple ones and, at the same time, give valuable insight into physical behavior. This chapter is concerned with some special techniques useful in circuit analysis.

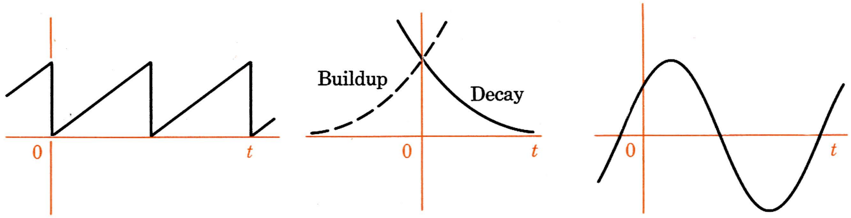
In analyzing circuits, the general procedure is to write element equations and connection equations and to solve them simultaneously. The voltages and currents used in transmitting all intelligence and most power are functions of time. Since the element equations include time derivatives and integrals, our first question might be: What time functions are important in engineering? Since the connection equations hold at every instant, our second question might be: How can continuously varying voltages and currents be added conveniently? In answering these questions, we shall first look at a variety of time functions or *signals*, study the exponential and the sinusoidal functions in detail, and then develop an efficient technique for handling sinusoids in practical problems.

Certain patterns of time variation or *waveforms* are of special significance in electrical engineering because they are encountered so frequently. Some of these are shown in Fig. 3.1.

A *direct* or *continuous* wave of voltage is supplied by a storage battery or a direct-current (dc) generator. A *step* of current flows when a switch is thrown, suddenly applying a direct voltage to a resistance; the current in the battery of a pocket radio is an example. A *pulse* of current flows if the switch is turned ON and then OFF; the flow of information in a computer consists of very short



(a) Continuous (b) Step (c) Pulse



(d) Sawtooth (e) Exponentials (f) Sinusoidal

Figure 3.1 Common signal waveforms.

pulses. A *sawtooth* wave increases linearly with time and then resets; this type of voltage variation causes the electron beam to move repeatedly across the screen of a television picture tube. A decaying *exponential* current flows if energy is stored in the electric field of a capacitor and allowed to leak off through a resistor; in an unstable system, a voltage may build up exponentially. A *sinusoidal* voltage is generated when a coil is rotated at a constant speed in a uniform magnetic field; also, oscillating circuits are frequently characterized by sinusoidal voltages and currents.

Exponentials

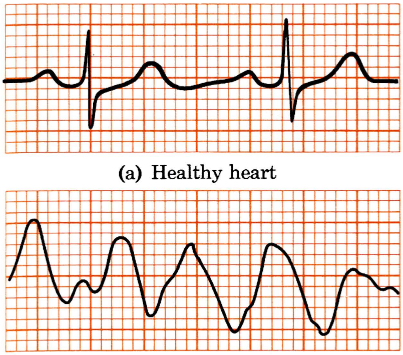
Exponential and sinusoidal waveforms are easy to generate and easy to analyze because of their simple derivatives and integrals; we shall pay them special attention because they are encountered frequently and also because they are useful in the analysis of more complicated waves, such as the electrocardiograms of Fig. 3.2.

The General Exponential

In general, an exponentially decaying quantity (Fig. 3.3) can be expressed as

$$a = A\,e^{-t/T} \tag{3-1}$$

where a = instantaneous value
A = amplitude or maximum value
e = base of natural logarithms = 2.718 . . .
T = time constant in seconds
t = time in seconds



(a) Healthy heart

(b) Heart in ventricular fibrillation

Figure 3.2 Electrocardiograms.

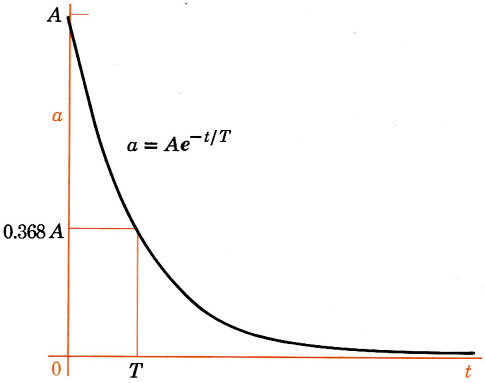


Figure 3.3 The decaying exponential.

The temperature of a hot body in cool surroundings, the angular velocity of a freely spinning bicycle wheel, and the current of a discharging capacitor can all be approximated by decaying exponential functions of time.

Time Constant

Since the exponential factor in Eq. 3-1 only *approaches* zero as t increases without limit, such functions theoretically last forever. In the same sense, all radioactive disintegrations last forever. To distinguish between different rates of disintegration physicists use the term "half-life," where the half-life of radium, say, is the time required for any amount of radium to be reduced to one half of its initial amount. In the case of an exponentially decaying current it is more convenient to use the value of time that makes the exponent -1. When $t = T =$ the *time constant,* the value of the exponential factor is

$$e^{-t/T} = e^{-1} = \frac{1}{e} = \frac{1}{2.718} = 0.368 \qquad (3\text{-}2)$$

In other words, after a time equal to the time constant, the exponential factor is reduced to approximately 37% of its initial value. (See Example 1, p. 64.)

Normalized Exponentials

All exponential equations can be reduced to a common form by expressing the variables in terms of dimensionless ratios or *normalizing* the variables. When this is done, Eq. 3-1 becomes

$$\frac{a}{A} = e^{-t/T} \qquad (3\text{-}3)$$

This equation leads to the *universal* exponential curve of Fig. 3.5. After 2 time

Example 1

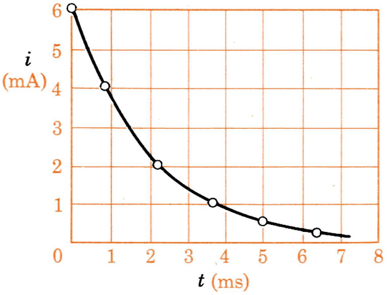
A capacitor is charged to a voltage of 6 V and then, at time $t = 0$, it is switched across a resistor and observations are made of current as a function of time. The experimental data are plotted in Fig. 3.4a. Draw the circuit, determine the time constant for this circuit, and derive an exponential equation for the current.

A smooth curve is drawn through the experimental points. The amplitude of the exponential is determined by noting that at $t = 0$, $i = I_0 = 6$ mA.

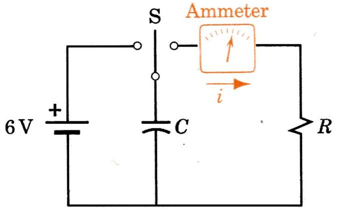
The time constant is determined by noting that when $i = I_0 e^{-1} = 6(0.368) \cong 2.2$ mA, $t = T \cong 2$ ms.

Then $1/T = 1/0.002 = 500$ and the current equation is

$$i = I_0 e^{-t/T} = 6 e^{-500t} \text{ mA} \tag{3-4}$$



(a) Data



(b) Circuit

Figure 3.4 Observation of time constant.

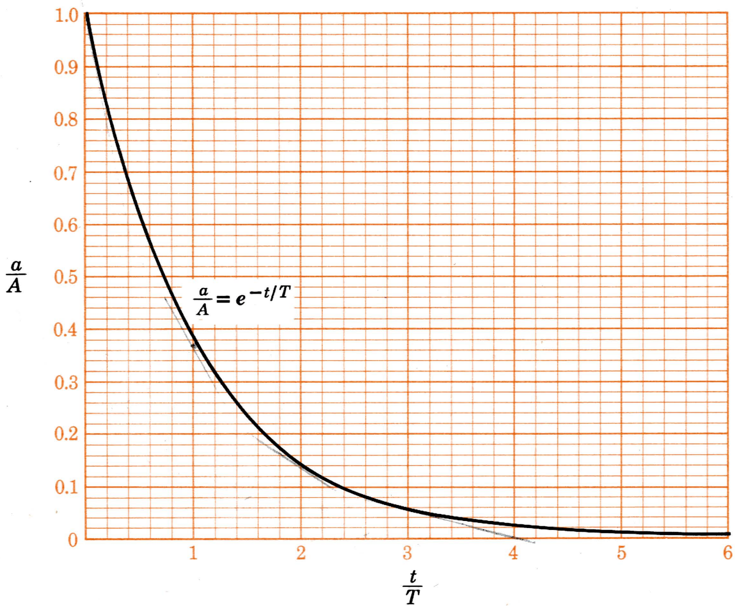


Figure 3.5 The universal exponential curve.

constants, $t = 2T$ and the function is down to $1/e^2$ times its initial value, or $a/A = (0.368)^2 = 0.135$. After 5 time constants, $t = 5T$ and the normalized function is down to $1/e^5 = 0.0067$. This value is less than the error expected in many electrical measurements, and it may be assumed that after 5 time constants an exponential current is practically negligible.

When plotted on semilog graph paper, an exponential function becomes a straight line with slope equal to the reciprocal of the time constant. As another interpretation of the time constant, it can be demonstrated (see Problem 1) that, if the exponential decay continued at its *initial rate*, the duration of the function would be just equal to T.

Exponential Voltages and Currents

Exponentials are important because many natural phenomena follow an exponential time variation. Also, the simplicity of the mathematical relations between exponentials and their derivatives and integrals makes them particularly valuable in circuit analysis, as in Example 2.

Example 2

In Fig. 3.6a the current $i = 5\,e^{-2t}$ A.
(a) Determine the voltage $v(t)$.

$$v_R = Ri = 4 \times 5\,e^{-2t} = 20\,e^{-2t}\ \text{V}$$

$$v_L = L\frac{di}{dt} = 1(-2)5\,e^{-2t} = -10\,e^{-2t}\ \text{V}$$

$$v = v_R + v_L = 20\,e^{-2t} - 10\,e^{-2t} = +10\,e^{-2t}\ \text{V}$$

(b) If the current $i = 5\,e^{-2t}$ A flows in an initially uncharged 0.25-F capacitance ($v_C = 0$ at $t = 0$ in Fig. 3.6b), determine the voltage $v_C(t)$.

$$v_C(t) = \frac{1}{C}\int_0^t i\,dt + V_0 = \frac{1}{0.25}\int_0^t 5\,e^{-2t}\,dt + 0$$

$$= \frac{5}{0.25(-2)}e^{-2t}\Big]_0^t = 10 - 10\,e^{-2t}\ \text{V}$$

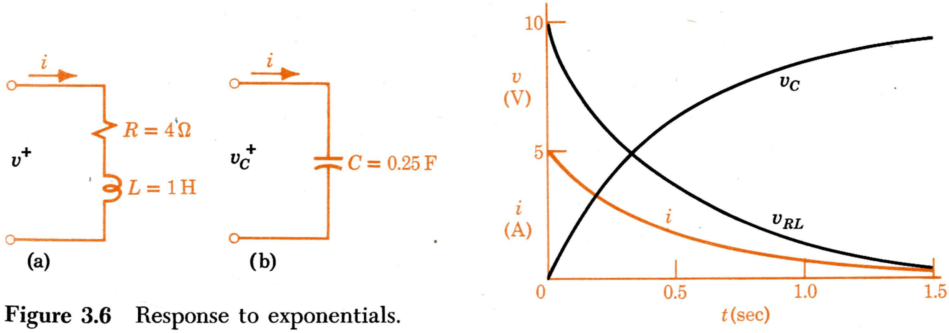


Figure 3.6 Response to exponentials.

Sinusoids

There are three principal reasons for emphasis on sinusoidal functions of time. First, many natural phenomena are sinusoidal in character; the vibration of a guitar string, the projection of a satellite on the rotating earth, and the current in an oscillating circuit are examples. Second, sinusoids are important in the generation and transmission of electric power and in the communication of intelligence; since the derivatives and integrals of sinusoids are themselves sinusoidal, a sinusoidal source always produces sinusoidal responses in any linear circuit. Third, other periodic waves can be represented by a series of sinusoidal components by means of Fourier analysis†; this means that techniques for handling sinusoids are useful in predicting the behavior of circuits involving pulses and sawtooth waves as well. Clearly, an efficient technique for analyzing circuits involving sinusoidal voltages and currents is essential. Since the application of Kirchhoff's laws involves summations of currents and voltages, we need a convenient method for adding quantities that are continually varying functions of time.

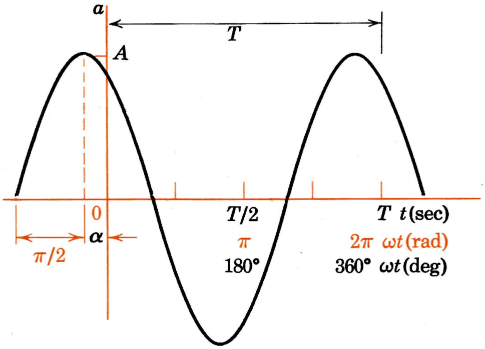


Figure 3.7 The general sinusoid.

The General Sinusoid

In general, a sinusoidally varying quantity (Fig. 3.7) can be expressed as

$$a = A \cos(\omega t + \alpha) \tag{3-5}$$

where

a = instantaneous value
A = amplitude or maximum value
ω = frequency in radians per second (omega)
t = time in seconds
α = phase angle in radians (alpha)

Since an angle of 2π radians corresponds to one complete cycle,

$$f = \frac{\omega}{2\pi} = \textit{frequency} \text{ in cycles per second or hertz (Hz)} \tag{3-6}$$

The time for one complete cycle is

$$T = \frac{1}{f} = \textit{period} \text{ in seconds} \tag{3-7}$$

The same sinusoid can be expressed by the sine-wave function

$$a = A \sin\left(\omega t + \alpha + \frac{\pi}{2}\right) \tag{3-8}$$

---

† See Skilling, *Electrical Engineering Circuits*, 2nd ed., John Wiley & Sons, New York, 1965, Chapter 14.

since sin $(x + \pi/2) = \cos x$ for all values of x. In either case, the instantaneous value varies from positive maximum to zero to negative maximum to zero repeatedly; a current that varies sinusoidally is called an *alternating current.*

Example 3

A sinusoidal current with a frequency of 60 Hz reaches a positive maximum of 20 A at $t = 2$ ms (Fig. 3.8). Write the equation of current as a function of time.

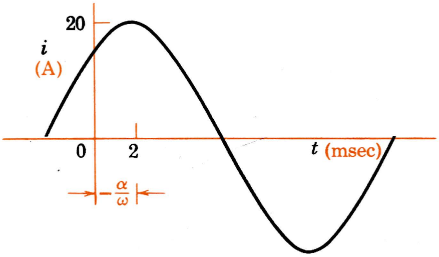


Figure 3.8 A sinusoidal current.

In the form of Eq. 3-5, $A = 20$ A and $\omega = 2\pi f = 2\pi(60) = 377$ rad/s. The positive maximum is reached when $(\omega t + \alpha) = 0$ or any multiple of 2π. Letting $(\omega t + \alpha) = 0$,

$$\alpha = -\omega t = -377 \times 2 \times 10^{-3} = -0.754 \text{ rad}$$

or

$$\alpha = -0.754 \times \frac{360°}{2\pi} = -43.2°$$

Therefore, the current equation is

$$i = 20 \cos (377t - 43.2°) \text{ A}$$

Note that the angle in parentheses is a convenient hybrid; to avoid confusion, the degree symbol is essential.

Sometimes it is convenient to associate the sinusoidal variation with the projection of a radial line of length A rotating with frequency ω from an initial or phase angle α. For radial line A of Fig. 3.9, the horizontal projection is $a = A \cos (\omega t + \alpha)$. The vertical projection of radial line A' is $a' = A' \sin (\omega t + \alpha + \pi/2)$, which is just equal to a; clearly, either a horizontal projection or a vertical projection can be used.

In a linear circuit consisting of resistances, inductances, and capacitances, if any voltage or current is a sinusoid, all voltages and currents will be sinusoids of the same frequency. To perform the addition and subtraction of voltages and currents required in the application of Kirchhoff's laws, three procedures are possible. In *graphical* addition, sinusoids are plotted and the curves added point by point; this method is time-consuming and lacks the generality needed to arrive at comprehensive solutions. The use of *trigonometric identities* permits the addition of subtraction of functions of time in general form, but this approach is laborious at best (see Exercise 18). The convenient and general method that we shall use is based on complex algebra and an ingenious transformation of functions of time into constant quantities.

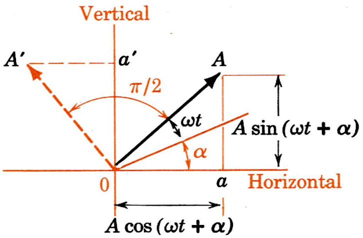


Figure 3.9 Projections of a rotating line.

Imaginary Numbers

A brief review may be worthwhile. In our numbering system, *positive* numbers correspond to distances measured along a line, starting from an origin. *Negative* numbers enable us to solve equations like $x + 2 = 0$, and they are represented in Fig. 3.10 by distances to the left of the origin. *Zero* is a relatively recent concept with some peculiar characteristics; for example, division by zero must be handled carefully. Numbers corresponding to distances along the line of Fig. 3.10 are called *real* numbers. To solve equations like $x^2 + 4 = 0$, so-called *imaginary* numbers were invented; these add a new dimension to our numbering system.

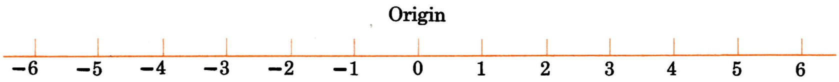


Figure 3.10 The system of real numbers.

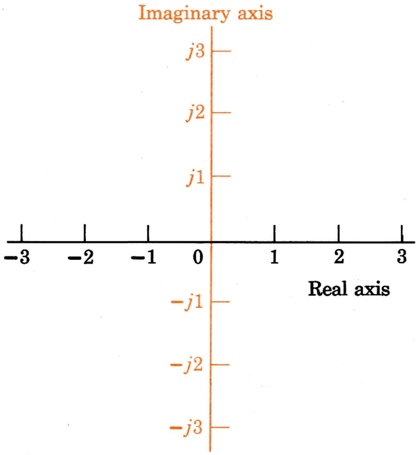


Figure 3.11 The complex plane.

Imaginary numbers are plotted along the *imaginary axis* of Fig. 3.11 so that $j2$[†] lies at a distance of 2 units along an axis at right angles to the *real axis*. The imaginary number $-j2$ lies along the imaginary axis but in the opposite direction from the origin. Note that all numbers are abstractions, and calling one number "real" and another "imaginary" just differentiates between their mathematical properties.

It may be helpful to think of j as an *operator*. An operator like "$\sqrt{\ }$" or "sin" means a particular mathematical operation; $\sqrt{\ }$ operating on 4 yields 2 and sin operating on 30° yields 0.5. In the same way, the minus sign $(-)$ is an operator such that $(-)$ operating on a directed line segment yields a reversal or a 180° rotation. Similarly, we define the operator j to mean a 90° counterclockwise rotation. The operator j taken twice, or $(j)(j)$, indicates a 180° rotation or a reversal, and $(j)(j)$ is equivalent to $(-)$ in effect. This concept is important because it provides an algebraic interpretation for a geometrical operation.

[†] Engineers use j instead of i for imaginary numbers to avoid confusion with the symbol for current. Charles P. Steinmetz, who first used combinations of real and imaginary numbers in circuit analysis, called them "general numbers." See *IEEE Spectrum*, April, 1965.

Just as $\sin^2 \theta$ (theta) is defined to mean $(\sin \theta)^2$ and not $\sin(\sin \theta)$, so the properties of jb must be defined. It is understood that

$$\begin{aligned} j(jb) &= j^2 b = -b & &\text{or we say} & j^2 &= -1 \\ j(j^2 b) &= j^3 b = -jb & &\text{or we say} & j^3 &= -j \\ j(j^3 b) &= j^4 b = +b & &\text{or we say} & j^4 &= +1 \end{aligned}$$

The equation $j^2 = -1$ is a statement that the two *operations* are equivalent and does *not* imply that j is a number. However, and this is the beauty of the concept, in all algebraic computations imaginary numbers can be handled as if j had a numerical value of $\sqrt{-1}$.

Complex Numbers

The real and imaginary axes define the *complex plane*. The combination of a real number and an imaginary number defines a point in the complex plane and also defines a *complex number*. The complex number may be considered to be the point or the directed line segment to the point; both interpretations are useful.

Given the complex number **W** of magnitude M and direction θ, in *rectangular* form,

$$\mathbf{W} = a + jb$$

or

$$\mathbf{W} = M(\cos\theta + j\sin\theta)$$

by inspection of Fig. 3.12. By Euler's theorem,†

$$\cos\theta + j\sin\theta = e^{j\theta} \qquad (3\text{-}9)$$

Therefore,

$$\mathbf{W} = M e^{j\theta} \qquad (3\text{-}10)$$

This is called the *exponential* or *polar* form and can be written symbolically as

$$\mathbf{W} = M \angle\theta \qquad (3\text{-}11)$$

which is read "magnitude M at angle θ." (In this book complex numbers are always in boldface type.)

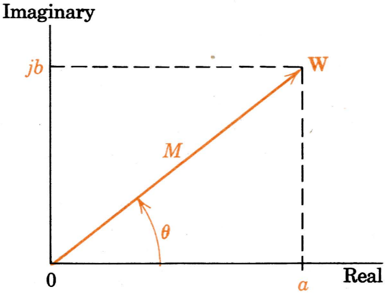


Figure 3.12 A complex number.

---

† The validity of Euler's theorem is evident when series expansions are written as follows:

$$e^\theta = 1 + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{\theta^5}{5!} + \cdots$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots \quad \text{and} \quad \sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots$$

Then

$$\cos\theta + j\sin\theta = 1 + j\theta - \frac{\theta^2}{2!} - j\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + j\frac{\theta^5}{5!} - \cdots = e^{j\theta}$$

To convert complex numbers to *polar* form from rectangular,

$$M = \sqrt{a^2 + b^2} \qquad \theta = \arctan\frac{b}{a} \tag{3-12}$$

To convert complex numbers to *rectangular* form from polar,

$$a = M\cos\theta \qquad b = M\sin\theta \tag{3-13}$$

Phasors

The complex constant $\mathbf{W} = M e^{j\theta}$ is represented by a fixed radial line. If the line rotates at an angular velocity ω, as shown in Fig. 3.13, $\mathbf{W}$ is a complex function of time and

$$\mathbf{W}(t) = M e^{j(\omega t+\theta)} \tag{3-14}$$

The projection of this line on the real axis is

$$W_{\text{real}} = f_1(t) = M\cos(\omega t + \theta)$$

and the projection on the imaginary axis is

$$W_{\text{imag}} = f_2(t) = M\sin(\omega t + \theta)$$

Either component could be used to represent a sinusoidal quantity; in this book we shall always work with the real components.

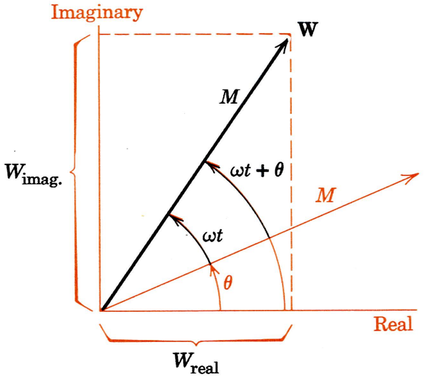


Figure 3.13 A complex function of time.

Phasor Representation

If an instantaneous voltage is described by a sinusoidal function of time such as

$$v(t) = V\cos(\omega t + \theta) \tag{3-15}$$

then $v(t)$ can be interpreted as the "real part of" a complex function or

$$v(t) = \text{Re}\,\{V e^{j(\omega t+\theta)}\} = \text{Re}\,\{(V e^{j\theta})(e^{j\omega t})\} \tag{3-16}$$

In the second form of Eq. 3-16, the complex function in braces is separated into two parts; the first is a complex constant, the second is a function of time that implies rotation in the complex plane. The first part we define as the *phasor* $\mathbf{V}$ where

$$\mathbf{V} = V e^{j\theta} = V\underline{/\theta} \tag{3-17}$$

Note that the phasor $V e^{j\theta}$ contains the important information: amplitude and phase angle. The term $e^{j\omega t}$ indicates rotation at angular velocity ω, but this is the same for all voltages and currents associated with a given source and therefore this term may be put to one side until needed.

Example 4

(a) Write the equation of the current shown in Fig. 3.14a as a function of time and represent the current by a phasor.

The current reaches a positive maximum of 10 A at $\pi/6$ rad or 30° before $\omega t = 0$; therefore the phase angle θ is $+\pi/6$ rad and the equation is

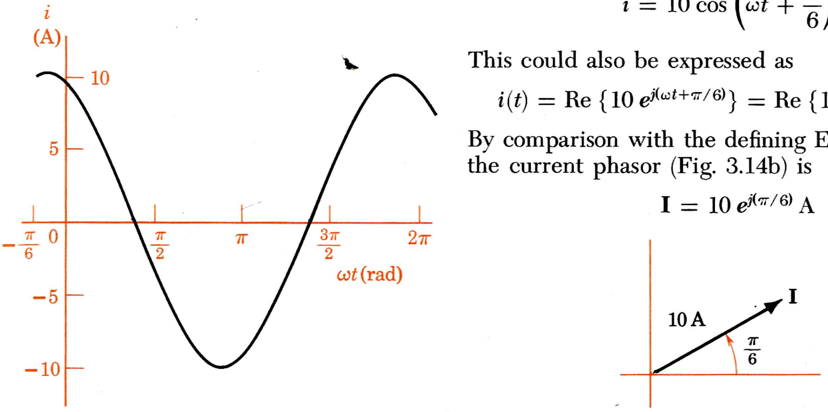
$$i = 10 \cos\left(\omega t + \frac{\pi}{6}\right) \text{ A}$$

This could also be expressed as

$$i(t) = \text{Re}\,\{10\, e^{j(\omega t + \pi/6)}\} = \text{Re}\,\{10\, e^{j(\pi/6)}\, e^{j\omega t}\} \text{ A}$$

By comparison with the defining Eqs. 3.16 and 3.17, the current phasor (Fig. 3.14b) is

$$\mathbf{I} = 10\, e^{j(\pi/6)} \text{ A}$$



(a) Time function (b) Phasor

Figure 3.14 Phasor representation.

(b) Represent, by a phasor, the current $i = 20 \sin(\omega t + \pi/6)$ A.

Our definition of phasor is based on the cosine function as the real part of the complex function of time. Since $\cos(x - \pi/2) = \sin x$, we first write

$$i = 20 \cos\left(\omega t + \frac{\pi}{6} - \frac{\pi}{2}\right) = 20 \cos\left(\omega t - \frac{\pi}{3}\right) \text{ A}$$

Then the current phasor is $\mathbf{I} = 20\, e^{j(-\pi/3)}$ A

The phasor **V** is called a *transform*† of the voltage $v(t)$; it is obtained by transforming a function of time into a complex constant that retains the essential information.

Complex Algebra

The advantage of representing sinusoids by phasors is that algebraic processes are simplified by using complex constants in place of trigonometric functions of time. Operations such as addition and subtraction, multiplication and division, raising to powers, and extracting roots are performed easily if the complex numbers are in the most convenient form. The following examples indicate the rules of complex algebra as applied to the two complex numbers

---

† See Skilling, *Electrical Engineering Circuits,* 2nd ed., John Wiley & Sons, New York, 1965, p. 145.

$$\mathbf{A} = a + jb = A\,e^{j\alpha} = A\,\underline{/\alpha} \qquad \text{(read: "}A\text{ at angle alpha")}$$

and

$$\mathbf{C} = c + jd = C\,e^{j\gamma} = C\,\underline{/\gamma} \qquad \text{(read: "}C\text{ at angle gamma")}$$

Equality

Two complex numbers are equal if and only if the real parts are equal and the imaginary parts are equal.

$$\text{If } \mathbf{A} = \mathbf{C}\text{, then } a = c \quad \text{and} \quad b = d \tag{3-18}$$

Example 5

The voltage phasors across two branches in parallel (Fig. 3.15) are $\mathbf{V}_1 = 2 + jy$ and $\mathbf{V}_2 = V\,\underline{/60°}$. Find y and V.

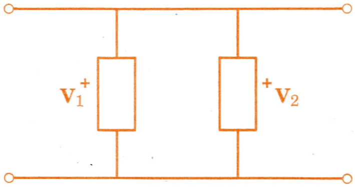


Figure 3.15 Phasor equality.

For branches in parallel, the voltages are equal and therefore the voltage phasors must be equal; hence,

$$2 + jy = V\,\underline{/60°}$$

This appears to be a single equation with two unknowns, but applying the rule for equality,

$$2 = V\cos 60° \quad \text{and} \quad y = V\sin 60°$$

Therefore,

$$V = \frac{2}{\cos 60°} = 4 \quad \text{and} \quad y = 4\sin 60° = 2\sqrt{3}$$

Addition

Two complex numbers in rectangular form are added by adding the real parts and the imaginary parts separately.

$$\mathbf{A} + \mathbf{C} = (a + c) + j(b + d) \tag{3-19}$$

Subtraction can be considered as addition of the negative,

$$\mathbf{A} - \mathbf{C} = (a - c) + j(b - d) \tag{3-20}$$

Addition and subtraction are not convenient in the polar form.

Example 6

In Fig. 3.16, $i_2(t) = 2\cos(\omega t - 90^\circ)$ and $\mathbf{I}_1 = 3 - j4$. Find $i_3(t)$.

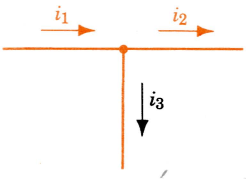


Figure 3.16 Phasor subtraction.

By inspection we write $\mathbf{I}_2 = 2\underline{/-90^\circ}$, but since the polar form is not convenient we convert to rectangular form as $\mathbf{I}_2 = 0 - j2$. Then

$$\begin{aligned}\mathbf{I}_3 &= \mathbf{I}_1 - \mathbf{I}_2 = (3 - j4) - (0 - j2)\\ &= 3 - j4 + j2 = 3 - j2\end{aligned}$$

In polar form, by Eq. 3.12,

$$\mathbf{I}_3 = \sqrt{3^2 + 2^2}\underline{/\arctan \tfrac{-2}{+3}} = 3.6\underline{/-33.7^\circ}$$

$$\therefore\ i_3(t) = 3.6\cos(\omega t - 33.7^\circ)$$

Multiplication

The magnitude of the product is the product of the individual magnitudes, and the angle of the product is the sum of the individual angles.

$$(\mathbf{A})(\mathbf{C}) = (A\,e^{j\alpha})(C\,e^{j\gamma}) = AC\,e^{j(\alpha+\gamma)} = AC\,\underline{/\alpha + \gamma} \tag{3-21}$$

Multiplication of complex numbers in polar form follows the law of exponents. If the numbers are given in rectangular form and the product is desired in rectangular form, it may be more convenient to perform the multiplication directly, employing the concept that $j = \sqrt{-1}$ and $j^2 = -1$.

$$\begin{aligned}(\mathbf{A})(\mathbf{C}) &= (a + jb)(c + jd)\\ &= ac + j^2bd + jbc + ajd = (ac - bd) + j(bc + ad)\end{aligned} \tag{3-22}$$

Example 7

In a certain alternating-current circuit $\mathbf{V} = \mathbf{ZI}$, where

$$\mathbf{Z} = 7.07\underline{/-45^\circ} = 5 - j5$$

and

$$\mathbf{I} = 10\underline{/+90^\circ} = 0 + j10.$$

Find $\mathbf{V}$.

In polar form,

$$\begin{aligned}\mathbf{V} &= (7.07\underline{/-45^\circ})(10\underline{/+90^\circ})\\ &= 7.07 \times 10\underline{/-45^\circ + 90^\circ} = 70.7\underline{/+45^\circ}\end{aligned}$$

In rectangular form,

$$\begin{aligned}\mathbf{V} &= (5 - j5)(0 + j10)\\ &= (5 \times 0 + 5 \times 10) + j(-5 \times 0 + 5 \times 10)\\ &= 50 + j50\end{aligned}$$

Division

The magnitude of the quotient is the quotient of the magnitudes, and the angle of the quotient is the difference of the angles.

$$\frac{\mathbf{A}}{\mathbf{C}} = \frac{A\,e^{j\alpha}}{C\,e^{j\gamma}} = \frac{A\underline{/\alpha}}{C\underline{/\gamma}} = \frac{A}{C}\underline{/\alpha - \gamma} \tag{3-23}$$

Division of complex numbers in polar form follows the laws of exponents.

Division in rectangular form is inconvenient but possible by using the *complex conjugate* of the denominator. Given a complex number $\mathbf{C} = c + jd$, the complex conjugate of $\mathbf{C}$ is defined as $\mathbf{C}^* = c - jd$; that is, the sign of the imaginary part is reversed. (Here the asterisk means "the complex conjugate of . . . .") Multiplying the denominator by its complex conjugate *rationalizes* the denominator (i.e., converts it to a real number) and simplifies division. To preserve the value of the quotient, the numerator is multiplied by the same factor.

Example 8

In a certain alternating-current circuit, $\mathbf{V} = \mathbf{ZI}$ where

$$\mathbf{V} = 130\underline{/-67.4^\circ} = 50 - j120$$

and

$$\mathbf{Z} = 5\underline{/53.1^\circ} = 3 + j4$$

Find **I**.

In polar form,

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{130\underline{/-67.4^\circ}}{5\underline{/53.1^\circ}} = 26\underline{/-120.5^\circ}$$

In rectangular form,

$$\begin{aligned}\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} &= \frac{50 - j120}{3 + j4}\cdot\frac{3 - j4}{3 - j4} \\ &= \frac{(150 - 480) + j(-360 - 200)}{9 + 16} \\ &= -13.2 - j22.4 = 26\underline{/-120.5^\circ}\end{aligned}$$

Powers and Roots. Applying the laws of exponents to a complex number in polar form,

$$\mathbf{A}^n = (A\,e^{j\alpha})^n = A^n\,e^{jn\alpha} = A^n\underline{/n\alpha} \tag{3-24}$$

The same result is obtained from the rule for multiplication by recognizing that raising to the nth power is equal to multiplying a number by itself n times.

Extracting the nth root of a number is equivalent to raising the number to the $1/n$th power and the same rule is applicable:

$$\mathbf{A}^{1/n} = (A\,e^{j\alpha})^{1/n} = A^{1/n}\,e^{j\alpha/n} = A^{1/n}\underline{/\alpha/n} \tag{3-25}$$

The method for finding all n roots is illustrated in Example 9.

Example 9

Perform the operations indicated at the right. Recalling that the number of distinct roots is equal to the order of the root (if n is an integer), locate the other roots in (b).

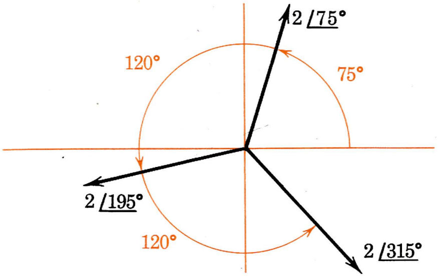


Figure 3.17 Roots of a complex number.

(a) $(8.66 - j5)^3 = (10\underline{/-30^\circ})^3$

$= 1000\underline{/-90^\circ} = 0 - j1000$

(b) $(-5.66 - j5.66)^{1/3} = (8\underline{/+225^\circ})^{1/3} = 2\underline{/+75^\circ}$

Since $8\underline{/225^\circ} = 8\underline{/225^\circ + N360^\circ}$ where N is any integer, three distinct roots are obtained by assigning values of $N = 0, 1, 2$.

For $N = 1$, $\quad (8\underline{/225^\circ + 360^\circ})^{1/3} = 2\underline{/75^\circ + 120^\circ}$

For $N = 2$, $\quad (8\underline{/225^\circ + 720^\circ})^{1/3} = 2\underline{/75^\circ + 240^\circ}$

The three cube roots are found to be complex numbers of the same magnitude but differing in angle by $360^\circ/3 = 120^\circ$ (see Fig. 3.17).

Phasor Diagrams

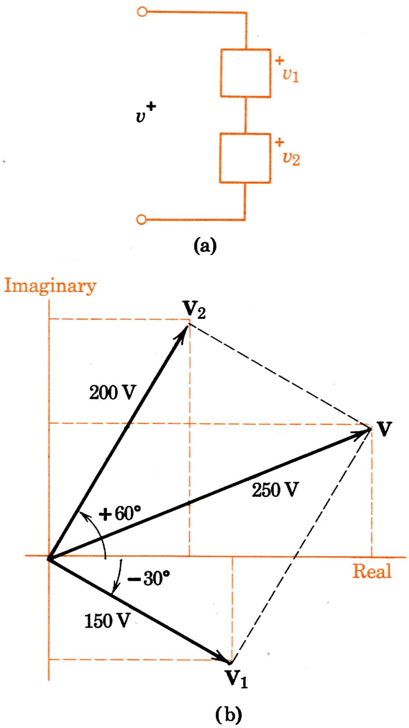
Electrical engineering deals with practical problems involving measurable currents and voltages that can exist in and across real circuit components such as those in your TV set. In solving circuit problems, however, we choose to deal with abstractions rather than physical currents and voltages. Learning to read and write and interpret such abstractions is an essential part of an engineering education. Working with the abstract concept that we call a phasor will provide some good experience in this important activity.

The phasor has much in common with the plane vector that is so valuable in solving problems in mechanics. As you recall, vector representation of forces permits the use of graphical methods for the composition and resolution of forces. Also, and this may be more important, the use of vector representation—in free-body diagrams, for example—provides a picture of the physical relationships that is not provided by the algebraic equations alone. In the same way, *phasor diagrams* can be used for graphical solution, as a quick check on the algebraic solution, and to gain new insight into voltage and current relations.

Example 10

Given $v_1 = 150\cos(377t - \pi/6)$ V and $\mathbf{V}_2 = 200\underline{/+60^\circ}$ V.

Find: $v = v_1 + v_2$.



(a)

(b)

Figure 3.18 Use of phasor diagrams.

The expression for v_1 is a mathematical abstraction carrying a great deal of information: a voltage varies sinusoidally with time, completing $377/2\pi = 60$ Hz and reaching a positive maximum value of 150 V when $t = (\pi/6)/377 = 0.00139$. To work with sinusoid v_1, we first *transform* the time function to phasor $\mathbf{V}_1 = 150\underline{/-30^\circ}$ V.

The expression for $\mathbf{V}_2$ is "an abstraction of an abstraction"; part of the information is specified and part is only implied. The sinusoidal character is implied by the boldface phasor notation. The maximum value is clearly 200 V and the phase angle is 60°. Since v_1 and v_2 are associated, we assume that they have the same angular frequency or $\omega_2 = 377$ rad/s. Reading "between the lines," we conclude that $v_2 = 200\cos(377t + \pi/3)$ V.

The phasor diagram of Fig. 3.18b shows the complex quantities $\mathbf{V}_1$ and $\mathbf{V}_2$ plotted to scale. Using the parallelogram law for adding vectors, the construction is as shown. Scaling off the resultant, it appears that $\mathbf{V}$ has a magnitude of about 250 V at an angle of approximately 23°.

For algebraic addition, the rectangular form is convenient, and we write (Eq. 3.13).

$$\mathbf{V}_1 = 150\cos(-30^\circ) + j150\sin(-30^\circ)$$
$$= 130 - j75\text{ V}$$
$$\mathbf{V}_2 = 200\cos 60^\circ + j200\sin 60^\circ = 100 + j173\text{ V}$$

By the rules for equality and addition,

$$\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2 = 230 + j98 = 250\underline{/23.1^\circ}\text{ V}$$

After the inverse transformation on $\mathbf{V}$,

$$v = v_1 + v_2 = 250\cos(377t + 23.1^\circ)\text{ V}$$

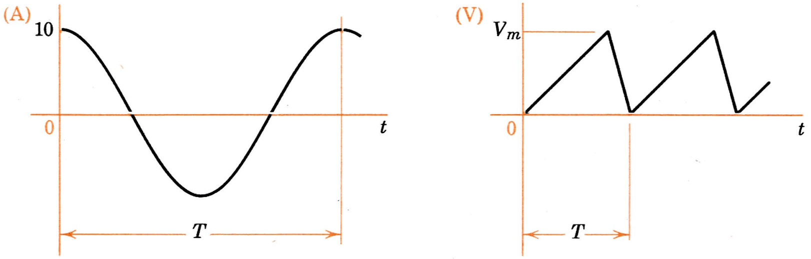
Looking at Fig. 3.18, we see that in Example 10 the algebraic addition involves the same components as the graphical addition, and we can conclude that phasor diagrams can be used for graphical calculations with the phasors treated just like vectors. A second virtue of the phasor diagram is that a quick sketch only approximately to scale provides a good check on algebraic calculations. Mistakes in sign, decimal point, or angle are easily spotted.

Periodic Waveforms

By definition, a *periodic* function of time is one in which

$$f(t + T) = f(t) \tag{3-26}$$

where T is the period in seconds. The sinusoidal current shown in Fig. 3.19a



(a) Sinusoidal (b) Triangular

Figure 3.19 Examples of periodic functions.

is periodic with a period of $T = 1/f = 2\pi/\omega$ seconds since

$$i = 10 \cos \omega \left(t + \frac{2\pi}{\omega}\right) = 10 \cos (\omega t + 2\pi) = 10 \cos \omega t$$

Always looking for ways to simplify calculations, we ask: What single value of current can be used to represent this continuously varying current?

The value to be used to represent a varying current depends on the function to be performed by the current. If it is used to actuate a relay, the maximum or *peak* value of 10 A is critical. If it is rectified (so that current flow is always in the same direction) and used to deposit silver in an electrolytic plating operation, the *average* value of 6.36 A is the significant quantity. If it is used to develop power in a resistor, the *effective* value of 7.07 A is the significant quantity. The term "peak value" is self-explanatory, but the other terms deserve consideration.

Average Value

The average value of a varying current $i(t)$ over the period T is the steady value of current I_{av}, which in period T would transfer the same charge Q. By definition,

$$I_{av}T = Q = \int_t^{t+T} i(t)\, dt = \int_0^T i\, dt$$

or

$$I_{av} = \frac{1}{T}\int_0^T i\, dt \tag{3-27}$$

In a similar way, average voltage is defined as

$$V_{av} = \frac{1}{T}\int_0^T v\, dt \tag{3-28}$$

In dealing with periodic waves, it is understood that the average is over one complete cycle (or an integral number of cycles) unless a different interval is specified. For example, the average value of any triangular wave (Fig. 3.19b) is equal to half the peak value since the area under the curve ($\int v\, dt$) for one cycle is $\frac{1}{2}V_mT$.

For a sinusoidal wave, the average value over a cycle is zero; the charge transferred during the negative half-cycle is just equal and opposite to that transferred during the positive half-cycle. In certain practical problems we are interested in the *half-cycle* average (Fig. 3.20) given by

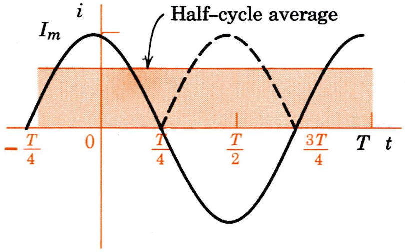


Figure 3.20 The half-cycle average.

$$
\begin{aligned}
I_{\text{half-cycle}} &= \frac{1}{\frac{1}{2}T}\int_{-T/4}^{+T/4} I_m \cos\frac{2\pi t}{T}\,dt \\
&= \frac{2I_m}{T}\left(\frac{T}{2\pi}\right)\sin\frac{2\pi t}{T}\Big]_{-T/4}^{+T/4} \\
&= \frac{I_m}{\pi}\left[\sin\left(\frac{\pi}{2}\right) - \sin\left(-\frac{\pi}{2}\right)\right] \\
&= \frac{2}{\pi}I_m = 0.636I_m
\end{aligned}
\tag{3-29}
$$

Effective Value

In many problems we are interested in the energy-transfer capability of an electric current. By definition, the average value of a varying power $p(t)$ is the steady value of power P_{av} that in period T would transfer the same energy W. If

$$
P_{\text{av}}T = W = \int_t^{t+T} p(t)\,dt = \int_0^T p\,dt
$$

then

$$
P_{\text{av}} = P = \frac{1}{T}\int_0^T p\,dt \tag{3-30}
$$

By convention, P always means average power and no subscript is necessary.

If electrical power is converted into heat in a resistance R,

$$
P = \frac{1}{T}\int_0^T p\,dt = \frac{1}{T}\int_0^T i^2R\,dt = I_{\text{eff}}^2R \tag{3-31}
$$

where I_{eff} is defined as the steady value of current that is equally *effective* in converting power. Solving Eq. 3-31,

$$
I_{\text{eff}} = \sqrt{\frac{1}{T}\int_0^T i^2\,dt} = I_{\text{rms}} \tag{3-32}
$$

and I_{eff} is seen to be the "square root of the mean squared value" or the *root-mean-square* current I_{rms}.

The effective or rms value of a sinusoidal current can be found from Eq. 3-32. Where $i = I_m \cos(2\pi/T)t$,

$$
I_{\text{rms}}^2 = \frac{1}{T}\int_0^T I_m{}^2\cos^2\frac{2\pi t}{T}\,dt = \frac{1}{T}\frac{I_m{}^2}{2}\int_0^T\left(1 + \cos\frac{4\pi t}{T}\right)dt = \frac{I_m{}^2}{2}
$$

Therefore, the effective value of a sinusoidal current is

$$I_{\text{rms}} = \frac{I_m}{\sqrt{2}} = 0.707 I_m \tag{3-33}$$

In Fig. 3.21, note that the mean value of $\cos^2 \omega t$ is just equal to 0.5 since the average of the $\cos 2\omega t$ term is zero over a full cycle. In words, Eq. 3-33 says: "For sinusoids, the effective value is $1/\sqrt{2}$ times the maximum value." For all other functions Eq. 3-32 must be used.

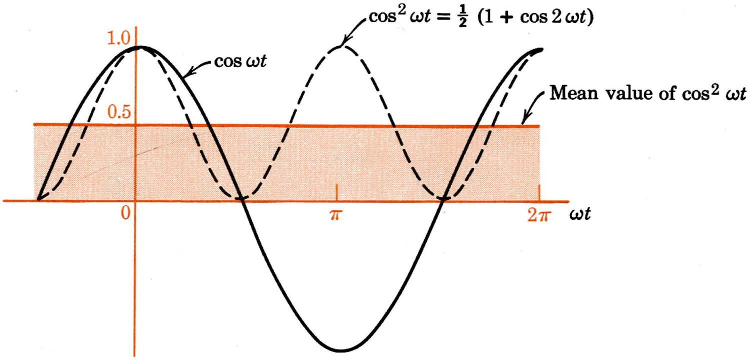


Figure 3.21 Calculating the root-mean-square value of a sinusoid.

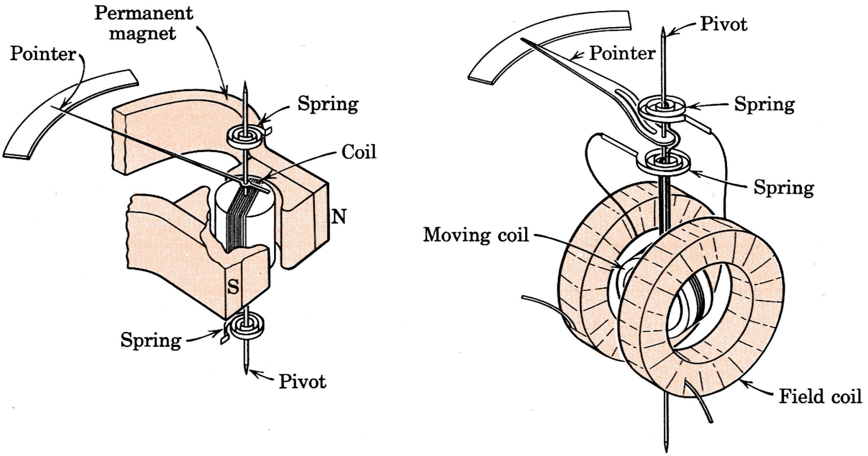
Meter Readings

Basically, a measuring instrument converts a physical effect into an observable quantity. The cathode-ray oscilloscope (CRO) converts an applied voltage into a deflection of the spot of impact of an electron beam. The beam has little inertia and will follow rapid voltage variations, and therefore the CRO indicates *instantaneous* values (see Figs. 10.5 and 10.9).

In the common dc ammeter (Fig. 3.22a), a current in the meter coil (suspended in a magnetic field) produces a torque opposed by a spiral spring. Although the torque is directly proportional to the instantaneous current, the high inertia of the movement (coil, support, and needle) prevents rapid acceleration and the observed deflection is proportional to the *average* current.

In a common type of ac ammeter (Fig. 3.22b), the magnetic field is set up by the current itself and the torque produced is proportional to the square of the current. Because of the inertia of the movement, the observed deflection is proportional to the *average squared* current and the scale is calibrated to indicate *rms* values.

If a periodic voltage is applied to a rectifier in series with a capacitance, the capacitance will tend to charge up until the applied voltage reaches its maximum.



(a) Direct current (b) Alternating current

Figure 3.22 Ammeter construction.

When the applied voltage decreases, the rectifier prevents a current reversal and the capacitance voltage remains at the *peak* value, which can be read by an appropriate dc instrument. The design and use of such instruments are described in Chapter 27, and this brief mention just emphasizes the practical nature of the values defined in this section.

Root-Mean-Square Rating

The root-mean-square value has another significance in specifying the "rating" of an electric motor for a varying duty cycle. The interpretation of "rating" depends on the type of device being rated. A fish line rated at "10-lb test" can be expected to fail at slightly over 10 lb of static pull. If an automobile engine is rated at "300 hp," it means that a well-tuned sample operated under ideal conditions free of most accessories develops an output close enough to 300 hp to satisfy the advertising department.† In contrast, a truck engine or an electric motor is usually rated at the power it can produce on a continuous basis over a long period of time without excessive deterioration.

Actually, most "50-hp" electric motors can develop two or three times this amount of power for short periods of time. If operated with an overload continuously, however, the excessive losses (proportional to i^2R) raise the operating temperature and the insulation deteriorates rapidly. If the rough approximation is made that current drawn is proportional to power output and heat generated is proportional to current-squared, the rms horsepower is the critical rating.

---

† One horsepower = 746 watts = 550 ft-lb/s. If a 3220-lb automobile actually had 300 hp available for acceleration it would accelerate from 0 to 60 mph in 2.4 s.

Example 11

An electric motor is to be specified for the duty cycle shown in Fig. 3.23. Full power of 15 hp is required for 2 min, the power decreases linearly for the next 3 min, the motor idles for 2 min, and then the cycle repeats.

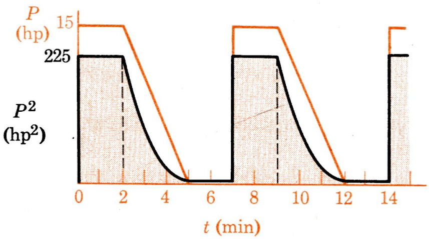


Figure 3.23 Rms horsepower calculation.

Assuming that the heat to be dissipated is proportional to the square of the power output, the problem is to find the rms horsepower. From Eq. 3.32,

$$\text{rms hp} = \sqrt{\frac{1}{T}\Sigma\,(\text{hp}^2 \times \text{time})} \qquad (3\text{-}34)$$

where the integral has been replaced by its equivalent, a summation. The form of Eq. 3.34 indicates a simple approach to this type of problem since Σ ($\text{hp}^2 \times$ time) is the area under the curve of hp^2 versus time. As shown in Fig. 3.23, the rectangular area is 225 $\text{hp}^2 \times 2$ min. Since the area under a parabola is $\frac{1}{3}$ that of the enclosing rectangle, the parabolic area is 225 $\text{hp}^2 \times 3$ min/3. Then the rms hp is

$$\sqrt{\tfrac{1}{7}(225 \times 2 + \tfrac{1}{3} \times 225 \times 3)} = \sqrt{96.4} = 9.8 \text{ hp}$$

The next larger standard rating is 10 hp and this rating would be specified.

Note that Example 11 is based on the assumption that the thermal capacity of the motor permits an averaging of the heating rate over a short period (7 minutes). If the duty cycle were in hours instead of minutes, a 15-hp motor would be required.

Effective Value Phasors

Where a sinusoidal voltage v is represented as the real part of the complex variable $V_m\, e^{j(\omega t + \theta)}$, the constant quantity $V_m\, e^{j\theta}$ has been called the phasor $\mathbf{V}$. The phasor carries information on the phase angle and maximum value of the sinusoid, but the value of importance in power calculations and the value read on ac meters is the effective value. Therefore, we now redefine $V_m\, e^{j\theta}$ as the phasor $\mathbf{V}_m$, and we use the phasor $\mathbf{V}$ to represent $\mathbf{V}_m/\sqrt{2}$, the *effective value phasor*. Diagrams drawn henceforth are identical with those drawn previously, except for an indiscernible change in voltage and current scales. Note that phasor $\mathbf{V}$ carries the same information as $\mathbf{V}_m$; given $\mathbf{V} = 10\underline{/30^\circ}$ V, now $v = 10\sqrt{2}\cos(\omega t + 30^\circ)$ V. Although this change may be confusing at first, the great benefit in problem solving will be clear later on.

SUMMARY

◆ The simplicity of the mathematical relations between exponentials and their derivatives and integrals makes them valuable in circuit analysis.

◆ The general decaying exponential is $a = A\,e^{-t/T}$.
The time constant T is a measure of the rate of decay.
When $t = T$, $a/A = 1/e = 0.368$; when $t = 5T$, $a/A = 1/e^5 = 0.0067$.

◆ Sinusoids are important because: they occur naturally, they are common in power and communication, they are easy to handle mathematically, and they can be used to represent other periodic waves.
The general sinusoid is $a = A \cos(\omega t + \alpha)$.
Frequency $\omega = 2\pi f$ rad/s; $f = 1/T$ hertz, where period T is in seconds.

◆ In the imaginary number jb, j is an operator indicating a 90° counterclockwise rotation; in algebraic manipulation, $j^2 1 = -1$ or $j = \sqrt{-1}$.
The combination of a real number a and an imaginary number jb defines a complex number $\mathbf{W} = a + jb = M\,e^{j\theta} = M\underline{/\theta}$. For conversion,

$$\begin{aligned} a &= M \cos\theta \qquad & M &= \sqrt{a^2 + b^2} \\ b &= M \sin\theta \qquad & \theta &= \arctan b/a \end{aligned}$$

◆ A sinusoidal function of time $a = A \cos(\omega t + \alpha)$ can be interpreted as the real part of the complex function $A\,e^{j(\omega t+\alpha)} = A\,e^{j\alpha} \cdot e^{j\omega t}$.
The complex constant $A\,e^{j\alpha}$ is defined as phasor $\mathbf{A}$, the transform of $a(t)$.
Phasor calculations follow the rules of complex algebra.

◆ Phasor diagrams follow the rules for vector diagrams and can provide an approximate answer, a quick check, or a clear visualization. The procedure is:

1. Transform the functions of time into constant phasors.
2. Perform the indicated operations on phasors, using phasor diagrams.
3. Transform the resulting phasors into functions of time.

◆ In a periodic function of time, $f(t + T) = f(t)$.

The average value of a periodic current is $I_{av} = \dfrac{1}{T}\displaystyle\int_0^T i\,dt$.

For a sinusoid, the half-cycle average is $(2/\pi)I_m = 0.636I_m$.

◆ The effective or rms value of a periodic current is $I = \sqrt{\dfrac{1}{T}\displaystyle\int_0^T i^2\,dt}$.

For a sinusoid, the rms value $I = I_m/\sqrt{2} = 0.707I_m$.
On phasor diagrams, use effective values as read on ac meters.

◆ Rms values are significant whenever the square of a variable is the important quantity. For function curves of simple geometry, the rms value is the square root of the average ordinate of the "squared" curve.

REVIEW QUESTIONS

1. Write the general equation of an exponentially *increasing* function.
2. Why is the "time constant" useful?
3. Justify the statement: "In a linear circuit consisting of resistances, inductances, and capacitances, if any voltage or current is an exponential all voltages and currents will be exponentials with the same time constant."
4. Write a statement for sinusoids similar to the foregoing statement for exponentials and justify it.
5. Write out in words a definition of a phasor.
6. How can a phasor, a constant quantity, represent a variable function of time?
7. Given $i_1 = 10 \cos (1000t + \pi/2)$ and $i_2 = 5 \cos 2000t$. Is the phasor method applicable to finding $i_1 + i_2$? Explain.
8. How are a quantity and its conjugate related on the complex plane?
9. Describe the quantity designated by the code symbols $\mathbf{I} = 20\underline{/\pi/4}$, $\omega = 60$.
10. Sketch three phasors on the complex plane and show the graphical construction to obtain $\mathbf{V}_1 + \mathbf{V}_2 - \mathbf{V}_3$.
11. Write an equation for $i(t)$ in terms of $\mathbf{I}$.
12. How can we neglect the imaginary part of the exponential function used in phasor representation?
13. Make up a table of trigonometric functions of 0, 30°, 45°, 60°, and 90° and for a 3–4–5 triangle; keep it handy for use.
14. Distinguish between average and effective values.
15. Why are ac ammeters calibrated to read rms values?
16. What is the reading of a dc ammeter carrying a current $i = 10 \cos 377t$ A?

EXERCISES

1. Sketch the following signals and write mathematical expressions for $v(t)$:
   (a) A step of 2 V occurring at $t = 1$ s.
   (b) A sawtooth of amplitude 50 V that resets 20 times a second.
   (c) A train of pulses of amplitude 100 V, duration 10 μs, and frequency 40 kHz.
   (d) An exponential voltage with an initial amplitude of 5 V that decays to 1 V in 80 ms. Predict its value at $t = 20$ ms.
2. An exponential voltage has a value of 15.5 V at $t = 1$ s and a value of 2.1 V at $t = 5$ s.
   (a) Derive the equation for $v(t)$.
   (b) How many time constants elapse between $t = 5$ s and $t = 9$ s?
   (c) Predict the value at $t = 9$ s.
3. An exponential current has a value of 8.7 mA at $t = 2$ s and a value of 3.2 mA at $t = 5$ s. Write the equation for $i(t)$ and predict the value at 11 s.
4. Throwing a switch in an experimental circuit develops a voltage $v = 6\,e^{-500t}$ V across a parallel combination of $R = 40\ \Omega$, and $C = 20\ \mu$F. Predict the current in each element and the total current.
5. Throwing a switch in an experimental circuit introduces a transient voltage $v = 120\,e^{-10t}$ V. If measurements in the circuit are accurate to 0.5%, approximately how long does the switching transient "last"?
6. A voltage $v = 12\,e^{-200t}$ V appears across a parallel combination of $R = 3$ kΩ, $C = 50\ \mu$F, and $L = 2$ H. Determine the current in each element and the total current.
7. Sketch, approximately to scale, the following functions:

   (a) $i_1 = 2 \cos (377t + \pi/4)$ A.
   (b) $v_1 = 10 \cos (300t - \pi/3)$ V.
   (c) $i_2 = 1.5 \sin (400t - 60°)$ A.

8. Express the following sinusoidal signals as specific functions of time:
   (a) A current with a frequency of 100 Hz and an amplitude of 10 mA and passing through zero with positive slope at $t = 1$ ms.
   (b) A voltage with a period of 20 μs and passing through a positive maximum of 5 V at $t = 5\ \mu$s.
   (c) A current reaching a positive maximum

of 2 A at $t = 5$ ms and the next negative maximum at $t = 25$ ms.

(d) A voltage having a positive maximum of 50 V at $t = 0$ and decreasing to a value of 25 V at $t = 2$ ms.

9. A sinusoidal signal has a value of -5 A at time $t = 0$ and reaches its first maximum (negative) of -10 A at $t = 2$ ms. Write the equation for $i(t)$ and predict the value at $t = 5$ ms. Determine the frequency and period.

10. A current $i = 2\cos(2000t - \pi/4)$ A flows through a series combination of $R = 20\ \Omega$ and $L = 10$ mH.

(a) Determine the voltage across each element.

(b) Sketch the two voltages, $v = f(\omega t)$, approximately to scale and estimate the amplitude and phase angle of $v_R + v_L$.

11. Given $\mathbf{A} = 3 + j4$, $\mathbf{B} = 2e^{j(\pi/6)}$, $\mathbf{C} = 4\angle{-90^\circ}$, sketch the three quantities on the complex plane and find:

(a) **AB**, (b) **B** − **C**, (c) **C**/**A**, (d) **CA***, (e) **B** + **AC**, (f) (**B** + **C**)/**A**, (g) $\mathbf{B}^2$ (h) $\sqrt{\mathbf{AC}}$, (i) $(\mathbf{C}/\mathbf{B})^{1/3}$

12. Given: $\mathbf{A} = 2\,e^{j(-\pi/3)}$, $\mathbf{B} = 1 + j1$, and $\mathbf{C} = 2\angle 90^\circ$, sketch the three quantities on the complex plane and find: (a) **AB**, (b) **B** − **C**; (c) **A**/**C**, (d) (**B** + **C**)**A**, (e) **CA***, (f) **B** + **AC**, (g) $\mathbf{B}^2$, (h) $\sqrt{\mathbf{AC}}$, (i) $(\mathbf{C}/\mathbf{B})^{1/3}$.

13. Find (a) the fourth roots of (-1), (b) the cube roots of $(-1 - j\sqrt{3})$.

14. Represent the time functions in Exercise 7 by phasors in polar and rectangular form.

15. Represent the signals in Exercise 8 by phasors in polar and rectangular form.

16. Write the time functions represented by the phasors given in Exercise 11.

17. In Fig. 3.24, the signals are $v_1 = 8\cos(\omega t - 30^\circ)$ V and $v_2 = 10\cos(\omega t + 45^\circ)$ V. Estimate **V** graphically from a phasor diagram drawn to scale and then calculate **V** from the component phasors. Determine $v(t)$.

18. In Fig. 3.25, $i_1 = 4\cos(\omega t + 30^\circ)$ and $i_2 = 6\cos(\omega t + 60^\circ)$.

(a) Determine **I** graphically from a phasor diagram drawn to scale and check by an analytical phasor solution, and determine $i(t)$.

(b) Expand the trigonometric expressions and determine $i(t)$ using trigonometric identities. Compare the difficulty of the two methods.

19. For each of the periodic voltage and current signals shown in Fig. 3.26,

(a) Calculate average values of voltage, current, and power.

(b) Calculate effective values of voltage and current.

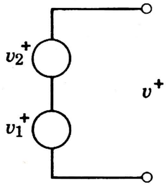


Figure 3.24

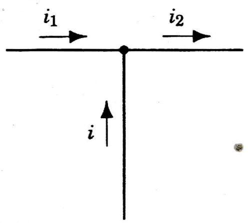
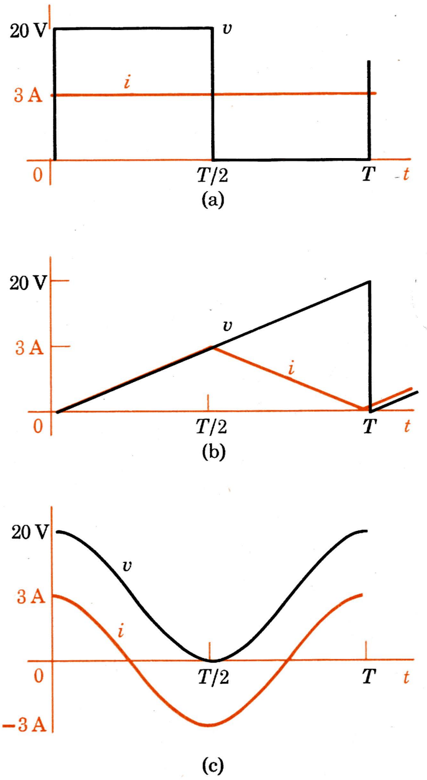


Figure 3.25



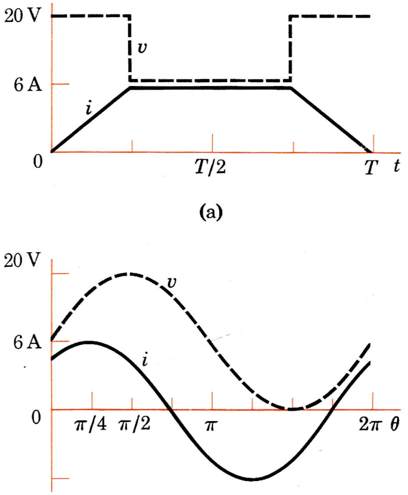
(a)

(b)

(c)

Figure 3.26

20. Calculate average and effective values of voltage and current for each of the periodic signals shown:
(a) In Fig. 3.27a.
(b) In Fig. 3.27b.



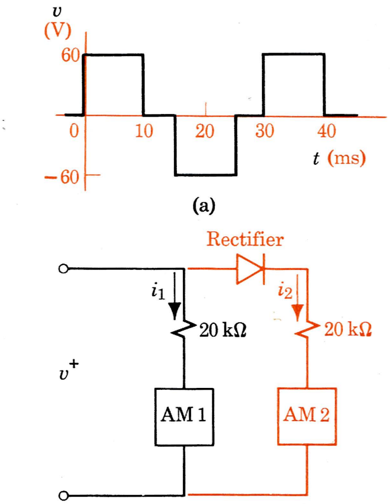
(a)

(b)

Figure 3.27

21. A signal consists of a series of positive rectangular pulses of magnitude V, duration T_d, and period T. Where $D = T_d/T$, derive expressions for average and effective values in terms of D and V.

22. The signal of Fig. 3.28a is applied to the circuit shown. The rectifier permits current flow only in the direction indicated by the triangle. Predict the readings on the two meters, if they are:
(a) dc ammeters,
(b) ac ammeters, of the types described in the text.



(a)

(b)

Figure 3.28

23. Repeat Exercise 22 if the voltage v is as shown in Fig. 3.29.

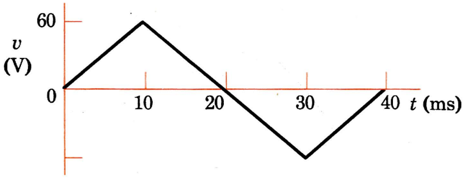


Figure 3.29

24. The duty cycle of an electric motor is: At $t = 0$ (motor idling and consuming negligible power), load increases abruptly to 60 hp and holds for 10 s, load decreases uniformly to 20 hp over 10 s and then is removed; motor idles for 40 s. Sketch the duty cycle (hp versus time) and specify the proper motor (available in 10-, 20-, 30-, 50-, and 100-hp ratings).

PROBLEMS

1. Demonstrate analytically that the tangent to any exponential function at time t intersects the time axis at $t + T$ where T is the time constant.

2. The voltage across a 10-μF capacitor is observed on an oscilloscope. In 0.5 s after a 10-V source is removed, the voltage has decayed to 1.35 V. Derive and label a circuit model for this capacitor.

3. When the rate of decrease of a quantity

(dq/dt) is directly proportional to the quantity itself (q), the quantity varies as an exponential function of time.

(a) Solve the differential equation $dq/dt = -kq$ by separating the variables, integrating both sides, and solving for $q(t)$.

(b) If q is electrical charge, draw an electrical circuit for which this differential equation is a mathematical model.

(c) From your solution to part (a), derive an expression for $i(t)$ in terms of the initial charge Q_0 and the circuit parameters.

4. Simple harmonic motion occurs when the acceleration d^2x/dt^2 is proportional to and oppositely directed to the displacement x. Write the differential equation governing simple harmonic motion and show that a sinusoidal displacement at angular velocity ω is a solution of the equation. Evaluate ω.

5. A voltage $v = 120\, e^{-2000t}$ V appears across a parallel combination of $R = 40\,\Omega$ and $C = 10\,\mu$F.

(a) Determine the total current through the combination as $f(t)$.

(b) Repeat part (a) for R increased to 50 Ω, and interpret this result physically.

6. A current $i = 2 \cos 2000t$ A flows through a series combination of $L = 30$ mH and $C = 5\,\mu$F. (C is initially uncharged.)

(a) Determine the total voltage across the combination as $f(t)$.

(b) Repeat part (a) for L increased to 50 mH and interpret this result physically.

7. A synchronous motor (described in Chapter 25) can be represented by the circuit model of Fig. 3.30 where $\mathbf{Z} = 10 + j30$ and $\mathbf{V} = \mathbf{E} + \mathbf{IZ}$. In each of the following cases, show the three quantities on a phasor diagram.

(a) For $\mathbf{V} = 2000\underline{/0^\circ}$ and $\mathbf{I} = 40\underline{/0^\circ}$, find **E**.

(b) For $\mathbf{E} = 2000\underline{/0^\circ}$ and $\mathbf{I} = 30\underline{/30^\circ}$, find **V**.

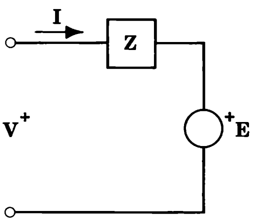


Figure 3.30

(c) For $\mathbf{V} = 2000\underline{/0^\circ}$ and $\mathbf{E} = 2000\underline{/-30^\circ}$, find **I**.

(d) For $\mathbf{V} = 2000\underline{/0^\circ}$ and $\mathbf{E} = 2000\underline{/+30^\circ}$, find **I**. Recalling that $p = vi$ and $w = \int vi\,dt$, determine whether under this condition the machine is absorbing electrical power as a motor or producing power as a generator.

8. Where h is defined as an operator producing a 120° counterclockwise rotation, express the following in polar and rectangular form (in terms of j): (a) $h1$, (b) $h^2 2$, (c) $(h2)^2$, (d) $(-h)2$, (e) $\sqrt{h4}$, (f) $4\sqrt{h^3}$.

9. Write the first six terms of the series expansion for e^x. Evaluate each term for $x = j1$. On the complex plane, plot each term to scale, adding onto the preceding term. Measure the sum of the six terms and express it in polar form. Is the result what you expected?

10. A voltage consists of a dc component of magnitude V_0 and a sinusoidal component of effective value V_1; show that the effective value of the combination is $(V_0^2 + V_1^2)^{1/2}$.

11. A voltage $v(t) = V + V \cos \omega t$ is applied across a series combination of R and L.

(a) Derive an expression for $i(t)$ in terms of the given quantities.

(b) Derive an expression for the average power dissipated in R.

12. A "three-phase generator" produces the voltages:

$v_{na} = 127\sqrt{2} \cos \omega t$ V,

$v_{ab} = 127\sqrt{2} \cos (\omega t - 120^\circ)$ V, and

$v_{nc} = 127\sqrt{2} \cos (\omega t - 240^\circ)$ V.

(a) Draw the three voltage phasors with n as a common node.

(b) Calculate v_{ab} and $\mathbf{V}_{ab}$.

(c) If generator terminals a, b, and c are connected to the corresponding terminals of the "load" in Fig. 3.31, predict the current in generator winding a.

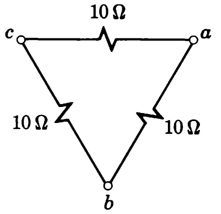


Figure 3.31

Chapter 4

Natural Response

First-Order Systems
Second-Order Systems
Impedance Concepts
Poles and Zeros

In predicting the behavior of electrical circuits or mechanical systems, we must take into account two different sources of energy. The behavior determined by an *external* energy source we call a *forced* response; the behavior due to *internal* energy storage we call a *natural* response. For example, a stretched string may be forced to vibrate at any frequency by an externally applied alternating force; if the string is plucked and released, however, it will vibrate at its characteristic frequency, its *natural* frequency, because of the internal energy stored in the plucking. As another example, an automobile may go fast or slow, forward or backward, as energy from the engine is controlled by the operator; but if the moving automobile is taken out of gear it will coast to rest in a predictable manner as the internal energy is dissipated.

In an electrical circuit, the external energy source or *forcing function* may be continuous or sinusoidal, or may have any of the other waveforms shown in Fig. 3.1. Energy may be stored internally in the electric field of a capacitor or the magnetic field of an inductor. In a chemical system, energy may be stored internally in the temperature of a liquid, in the potential energy of a compressed gas, or in the kinetic energy of a moving fluid. In the general problem both external and internal energy sources are present and both forced and natural responses must be considered.

Forced responses can be maintained indefinitely by the continual input of energy. In contrast, because of unavoidable dissipation in electrical resistance or mechanical friction, natural responses tend to die out. After the natural behavior has become negligibly small, conditions are said to have reached the *steady state*. The period beginning with the initiation of a natural response and ending when the natural response becomes negligible is called the *transient* period.

Although the natural behavior of a system is not itself a response to any external function, a study of natural behavior will reveal characteristics of a system that will be useful in predicting the forced response as well. Our approach is first to investigate the natural behavior, then to study the forced behavior, and finally to consider the general case of forced and natural behavior. In this chapter the emphasis is on the natural behavior of electrical circuits with some applications to mechanical and thermal systems. We start with relatively simple systems, derive the basic concepts, and then formulate a general procedure. The methods developed are powerful and they are applicable to problems in many branches of engineering.

Equations of First-Order Systems

In general, the behavior of an electrical circuit or any other physical system can be described by an integro-differential equation. In electrical circuits, the governing equation is obtained by applying the experimental laws for circuit elements and combinations of elements, taking into account external and internal energy sources.

In the circuit of Fig. 4.1, for example, a coil of inductance L and resistance R carrying a current I_0 is suddenly shorted by throwing switch S at time $t = 0$. The equation governing current i as a function of time t after the switch is thrown can be obtained by applying Kirchhoff's voltage law to the shorted series circuit. For all time $t > 0$, the sum of the voltages around the closed path (clockwise) must be zero or

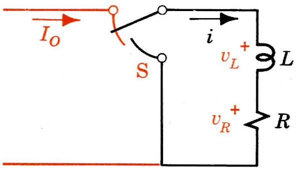


Figure 4.1 *LR* example.

$$\Sigma v = 0 = -v_L - v_R = -L\frac{di}{dt} - Ri$$

or

$$L\frac{di}{dt} + Ri = 0 \tag{4-1}$$

Note that there is no voltage across the short circuit even if current I_0 continues to flow.

As another example, in Fig. 4.2 capacitance C with initial voltage V_0 (or initial charge $Q_0 = CV_0$) is suddenly shorted across a resistance R at time $t = 0$. For all time $t > 0$,

$$\Sigma v = 0 = v_C - v_R = V_0 - \frac{1}{C}\int_0^t i\,dt - Ri$$

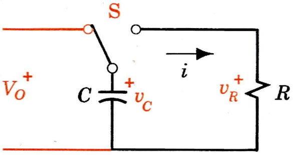


Figure 4.2 *RC* example.

Differentiating to eliminate the integral and the constant and rearranging the terms

$$R\frac{di}{dt} + \frac{1}{C}i = 0 \tag{4-2}$$

In Fig. 4.3, the mass M is sliding on a surface for which the coefficient of friction is D. The driving force is removed at $t = 0$ when the velocity is U_0 (or M is given an initial momentum MU_0). If the mass is allowed to coast, the only forces acting on M are those due to inertia and friction. Applying d'Alembert's principle, for all time $t > 0$,

$$\Sigma f = 0 = -M\frac{du}{dt} - Du$$

or

$$M\frac{du}{dt} + Du = 0 \tag{4-3}$$

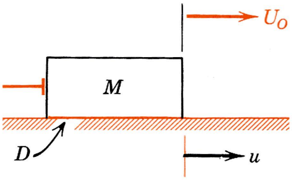


Figure 4.3 *MD* example.

These examples result in analogous, first-order, linear, homogeneous, ordinary differential equations—*analogous* because they are of the same mathematical form, *first-order* because that is the order of the highest derivative present, *linear* because the variable and its derivatives appear only to the first power, *homogeneous* because there are no constant terms or forcing functions, and *ordinary* because there is a single independent variable t.

The homogeneous equation contains only terms relating to the circuit itself; the solution of this equation is independent of any forcing function and therefore it is the natural response of the circuit. This type of equation occurs so frequently that we must learn a convenient method for solving it.

Solving the Equation

The solution of a differential equation consists in finding a function that satisfies the equation. Consider the equation for the circuit of Fig. 4.2,

$$R\frac{di}{dt} + \frac{1}{C}i = 0 \tag{4-2}$$

This equation says that the combination of a function $i(t)$ and its derivative di/dt must equal zero; the function and its derivative must cancel somehow. This implies that the form of the function and that of its derivative must be the same and suggests an exponential function. If we let

$$i = A\,e^{st} \qquad \text{then} \qquad \frac{di}{dt} = sA\,e^{st} \tag{4-4}$$

where A is an amplitude and s is a frequency.† Substituting these expressions into the homogeneous equation,

$$RsA\,e^{st} + \frac{1}{C}A\,e^{st} = 0$$

Therefore, for all values of time t,

---

† Note that the units of s are "per second." The reason for calling s a "frequency" is explained on p. 98.

$$\left(Rs + \frac{1}{C}\right) A\, e^{st} = 0 \tag{4-5}$$

In looking for solutions to this equation, $A = 0$ is one possibility, but this is the trivial case. The other possibility is to let $(Rs + 1/C) = 0$. This yields

$$s = -1/RC$$

and

$$i = A\, e^{(-1/RC)t} \tag{4-6}$$

To evaluate the amplitude A, we return to the original equation

$$V_O - \frac{1}{C}\int_0^t i\, dt - Ri = 0 \tag{4-7}$$

and consider a time $t = 0^+$, the first instant of time after $t = 0$. The integral of current over time represents the flow of charge and it takes a finite time for charge to be transferred. At $t = 0^+$, no time has elapsed and $\int_0^{0+} i\, dt = 0$. Therefore the initial conditions are that

$$V_O - 0 - Ri_O = 0 \qquad \text{or} \qquad i_O = \frac{V_O}{R} \tag{4-8}$$

By Eq. 4-6, at $t = 0^+$,

$$i_O = A\, e^{s(0)} = A\, e^0 = A$$

Therefore,

$$A = i_O = \frac{V_O}{R}$$

and for all values of time after the switch is closed

$$i = \frac{V_O}{R} e^{-t/RC} \tag{4-9}$$

The physical explanation of this mathematical expression is based on the fact that resistor current is proportional to voltage, whereas capacitor current is proportional to rate of change of voltage. Upon closing the switch, the full voltage of the capacitor is applied across the resistor and initial current is high $(i_O = V_O/R)$. But this current flow removes charge and reduces the voltage $(dv_C/dt = i/C)$ across the capacitor and resistor and therefore the current is reduced. As the voltage decreases, the current is reduced, the voltage changes more slowly, and the current becomes ever smaller.

As another approach to evaluating i_O, we recall that the energy stored in a capacitance is $\frac{1}{2}Cv^2$ and this energy cannot be changed instantaneously since that would involve an infinite power. Therefore, the voltage on a capacitance cannot be changed instantaneously and we conclude that the voltage just after switching (at $t = 0^+$) must equal the voltage just before switching, V_O. When the switch is closed, this voltage appears across the resistance and

$$v_C = V_O = v_R = i_O R$$

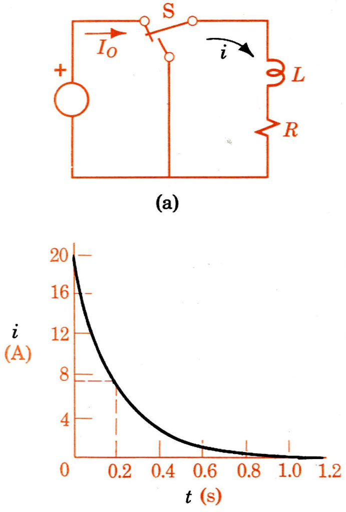
hence

$$i_O = \frac{V_O}{R}$$

as before. This approach is used in Example 1.

Example 1

Switch S (Fig. 4.4a) is arranged to disconnect the source and simultaneously, at time $t = 0$, short-circuit the coil of 2-H inductance and 10-Ω resistance. If the initial current in the coil I_O is 20 A, predict the current i after 0.2 s has elapsed.



(a)

(b)

Figure 4.4 Exponential current decay in an RL circuit.

How long will it take for the natural behavior current to become zero?

1. By Kirchhoff's voltage law,

$$\Sigma v = 0 = -v_L - v_R = -L\frac{di}{dt} - Ri$$

2. The homogeneous equation is

$$L\frac{di}{dt} + Ri = 0$$

3. Assuming an exponential solution, we write

$$i = A\,e^{st}$$

where s and A are to be determined.

4. Substituting into the homogeneous equation,

$$LsA\,e^{st} + RA\,e^{st} = (sL + R)A\,e^{st} = 0$$

If $sL + R = 0$, $s = -\dfrac{R}{L}$ and $i = A\,e^{-(R/L)t}$

5. The energy stored in an inductance, $\frac{1}{2}Li^2$, cannot change instantaneously. Therefore, the current in the coil just after the switch is thrown must equal the current just before. At $t = 0^+$,

$$i = I_O = A\,e^0 = A \quad \text{or} \quad A = I_O = 20$$

hence the solution is

$$i = I_O\,e^{-(R/L)t} = 20\,e^{-(10/2)t} = 20\,e^{-5t}\text{ A} \qquad (4\text{-}10)$$

As shown in Fig. 4.4b, after 0.2 s the current is

$$i = 20\,e^{-5\times0.2} = 20 \times 0.368 = 7.36\text{ A}$$

The current decreases continually but never becomes zero.

General Procedure

Analysis of the steps followed in Example 1 suggests a general procedure for determining the natural behavior of an electrical circuit:

1. Write the governing equation, using Kirchhoff's laws.
2. Reduce this to a homogeneous differential equation.
3. Assume an exponential solution with undetermined constants.
4. Determine the exponents from the homogeneous equation.
5. Evaluate the coefficients from the given conditions.

In a dynamic mechanical system, the first step usually involves the application of d'Alembert's principle and the equation for forces in equilibrium. Chemical engineering systems involving mass and energy transport may be analyzed in a similar way. Any such lumped linear system can be described by an ordinary differential equation that can be reduced to the homogeneous equation by eliminating constant terms and forcing functions. The solution of such an equation is always expressible in terms of exponential functions. The evaluation of the coefficients, which may be the most difficult step, is usually based on consideration of the energies stored in the system.

If current is the rate of transfer of charge and if the current in Example 1 flows "forever," will an infinite charge be transferred? To answer this question, express the total charge as the integral of current over an infinite time, or

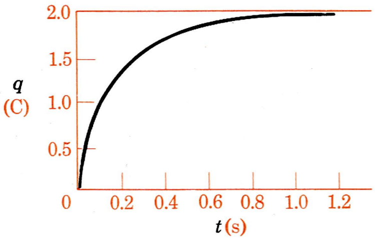


Figure 4.5 Total charge transfer.

$$q = \int_0^t i\,dt = \int_0^\infty 10\,e^{-5t}\,dt$$

$$= -\frac{10}{5}e^{-5t}\Big]_0^\infty$$

$$= -\frac{10}{5}(0 - 1) = 2\text{ C}$$

Although the current flows "forever," the rate of flow is decreasing in such a way that the total charge transferred approaches a finite limit (Fig. 4.5). The same is true of the energy dissipated in the resistance R (see Problem 4).

Time Constant

The term "first-order system" refers to a system involving only a single energy-storage element for which the governing equation is a first-order differential equation. The natural behavior obtained by solving the homogeneous equation will always be similar in form to the decay of current or voltage in an RL or RC circuit where

$$i = I_0\,e^{-t/T} \quad \text{or} \quad v = V_0\,e^{-t/T} \tag{4-11}$$

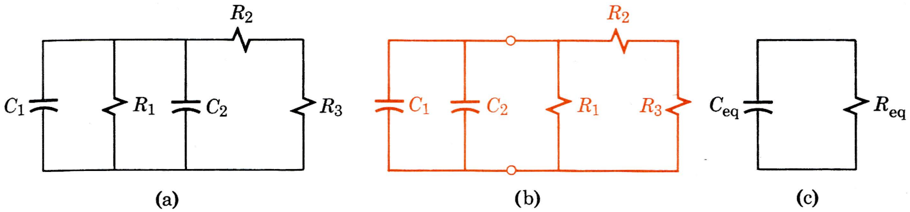


Figure 4.6 An *RC* circuit with a single time constant.

For the RL circuit of Example 1, the *time constant* is $T = L/R = 2/10 = 0.2$ s. For the RC circuit of Fig. 4.2, the current and the voltage reach $e^{-1} \doteq 0.368$ times the initial values at $T = RC$.

The circuit of Fig. 4.6 is a special case in that the two capacitances can be combined and replaced by an equivalent capacitance $C_{eq} = C_1 + C_2$. The three resistances can be combined and replaced by an equivalent resistance $R_{eq} = R_1(R_2 + R_3)/(R_1 + R_2 + R_3)$. The time constant for every voltage or current in this circuit is $T = R_{eq}C_{eq}$. (See Exercise 9a.)

Natural Response of Second-Order Systems

If there is more than one energy-storage element, the behavior of the system is more complicated but the analysis follows the same procedure as for the simpler systems. In an *RLC* circuit, energy may be stored in the inductance or in the capacitance. In the series circuit of Fig. 4.7, assume that an initial voltage V_0 exists on the capacitance *C*. With the switch S open there is no current in the inductance *L* and therefore no energy storage in its magnetic field. The resistance *R* is incapable of energy storage.

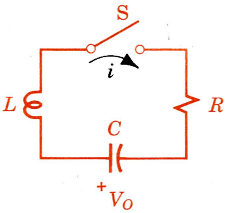


Figure 4.7 A series *RLC* circuit.

Following the general procedure for determining natural behavior, we first write the governing equation using Kirchhoff's voltage law. Traversing the loop (after the switch is closed) in a clockwise direction, we have

$$\Sigma v = 0 = -L\frac{di}{dt} - Ri + V_0 - \frac{1}{C}\int_0^t i\,dt$$

By differentiating and rearranging terms, the homogeneous equation is

$$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i = 0 \tag{4-12}$$

The presence of the second energy-storage element has increased the order of the highest derivative appearing in the homogeneous equation. This equation is typical of a *second-order* system.

The Characteristic Equation

Assuming an exponential solution, let $i = A\,e^{st}$. Substituting in the homogeneous equation yields

$$s^2LA\,e^{st} + sRA\,e^{st} + \frac{1}{C}A\,e^{st} = 0$$

which is satisfied when

$$Ls^2 + Rs + \frac{1}{C} = 0 \tag{4-13}$$

This equation is noteworthy. For this series circuit it is the sum of three terms, each relating to a circuit element. It contains no voltages or currents, only terms characteristic of the circuit. Also, from this equation we expect to obtain information about the character of the natural behavior; it is called the *characteristic equation* and is very useful in circuit analysis.

The roots of the characteristic equation (Eq. 4-13) are

$$s_1 = -\frac{R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}}$$

and

$$s_2 = -\frac{R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}} \tag{4-14}$$

If either $i_1 = A_1\,e^{s_1t}$ or $i_2 = A_2\,e^{s_2t}$ satisfies a linear homogeneous equation, i.e., makes it zero, then the sum of the two terms also satisfies the equation. The most general solution then is

$$i = A_1\,e^{s_1t} + A_2\,e^{s_2t} \tag{4-15}$$

where A_1 and A_2 are determined by the initial conditions and s_1 and s_2 are determined by the circuit constants.

The values of R, L, and C are all real and positive, but the values of s_1 and s_2 may be real, complex, or purely imaginary. The clue to the character of the natural response of a second-order system is found in the quantity under the radical sign—the *discriminant.* If the discriminant is positive, the roots of the characteristic equation will be real, negative, and distinct. If the discriminant is zero, the two roots will be real, negative, and identical. If the discriminant is negative, the roots will be complex numbers or, in the special case where $R = 0$, purely imaginary.

Roots Real and Distinct. If the discriminant in Eq. 4-14 is positive, s_1 and s_2 are real, negative, and distinct, and the natural behavior is the sum of two decaying exponential responses. Example 2 illustrates this case.

Example 2

Given the circuit values shown in Fig. 4.8a, determine and plot the current response as a function of time after the switch is closed.

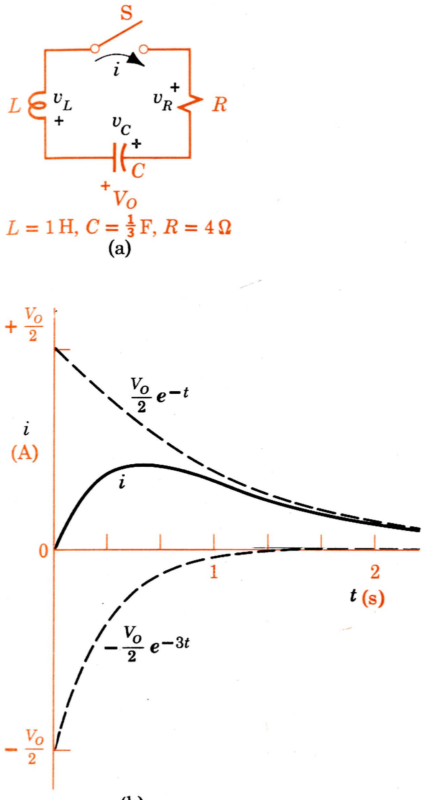


Figure 4.8 Natural response of an overdamped second-order system.

Roots of the characteristic equation (Eq. 4-14) are

$$s_1, s_2 = -\frac{4}{2\times 1} \pm \sqrt{\left(\frac{4}{2\times 1}\right)^2 - \frac{1}{1\times\frac{1}{3}}} = -2 \pm 1$$

The general solution for current (Eq. 4-15) is

$$i = A_1 e^{-t} + A_2 e^{-3t} \tag{4-16}$$

To evaluate the constants, consider the initial conditions. At the instant the switch was closed, the voltage across the capacitance $v_C = -V_O$ and the current in the inductance $i_L = 0$. Because energy cannot be changed instantaneously, these values must hold just after closing the switch, at $t = 0^+$, say.

At $t = 0^+$, the current in each element in this series circuit must be zero or (Eq. 4-16):

$$i = i_O = 0 = A_1 e^0 + A_2 e^0 = A_1 + A_2,$$

therefore $A_2 = -A_1$.

Also, the voltage iR across the resistance is zero. Therefore $v_L = -v_C = V_O$ and

$$L\frac{di}{dt} = V_O \qquad \text{or} \qquad \frac{di}{dt} = \frac{V_O}{L}$$

From Eq. 4-16 at $t = 0^+$

$$\frac{di}{dt} = \frac{V_O}{L} = -A_1 e^0 - 3A_2 e^0 = -A_1 - 3A_2$$

$$= -A_1 + 3A_1 = +2A_1$$

Solving,

$$A_1 = +\frac{V_O}{2L} = +\frac{V_O}{2} \quad \text{and} \quad A_2 = -A_1 = -\frac{V_O}{2}$$

The specific equation for the current response, plotted in Fig. 4.8b, becomes

$$i = \frac{V_O}{2}e^{-t} - \frac{V_O}{2}e^{-3t}$$

Roots Complex. If the discriminant in Eq. 4-14 is negative, s_1 and s_2 are complex conjugates and the natural behavior is the exponentially damped sinusoid defined by Eq. 4-17 in Example 3. However, the physical interpretation of Eq. 4-17 is not immediately clear. To gain insight into this type of response, let us consider

Example 3

In the circuit of Example 2, let $L = 1$ H, $C = \frac{1}{17}$ F and $R = 2\ \Omega$. Derive an expression for the natural response.

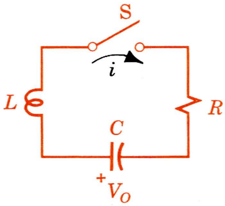


Figure 4.9 An *RLC* circuit with complex roots.

From Eq. 4-14, now

$$s = -\frac{R}{2L} \pm \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}}$$

$$= -\frac{2}{2\times 1} \pm \sqrt{\left(\frac{2}{2\times 1}\right)^2 - \frac{1}{1\times\frac{1}{17}}} = -1 \pm \sqrt{-16}$$

Therefore,

$$s_1 = -1 + j4 \qquad \text{and} \qquad s_2 = -1 - j4$$

and the natural response is

$$i = A_1\, e^{(-1+j4)t} + A_2\, e^{(-1-j4)t} \tag{4-17}$$

the case of complex roots in a general way. The mathematics will be simplified by arbitrarily defining three new terms:

$$\alpha = \frac{R}{2L} \qquad \omega_n{}^2 = \frac{1}{LC} \qquad \omega^2 = \frac{1}{LC} - \frac{R^2}{4L^2} = \omega_n{}^2 - \alpha^2 \tag{4-18}$$

Now, $\quad s_1 = -\alpha + j\omega \quad$ and $\quad s_2 = -\alpha - j\omega$

and the general form of Eq. 4-17 is:

$$i = A_1\, e^{(-\alpha+j\omega)t} + A_2\, e^{(-\alpha-j\omega)t} \tag{4-19}$$

By Euler's equation (Eq. 3-9),

$$e^{j\omega t} = \cos \omega t + j \sin \omega t \qquad \text{and} \qquad e^{-j\omega t} = \cos \omega t - j \sin \omega t$$

Factoring out the $e^{-\alpha t}$ and substituting for $e^{j\omega t}$ yields

$$i = e^{-\alpha t}[(A_1 + A_2)\cos \omega t + j(A_1 - A_2)\sin \omega t]$$

where A_1 and A_2 are constants that may be complex numbers. In fact, since these are physical currents, the coefficients of the cos ωt and sin ωt terms, $(A_1 + A_2)$ and $j(A_1 - A_2)$, must be real numbers that could be called B_1 and B_2 so that

$$i = e^{-\alpha t}[B_1 \cos \omega t + B_2 \sin \omega t] \tag{4-20}$$

But the sum of a cosine function and a sine function must be another sine function (properly displaced in phase), so Eq. 4-20 can be rewritten to give a natural response of

$$i = A\, e^{-\alpha t} \sin(\omega t + \theta) \tag{4-21}$$

This is evidently a sinusoidal function of time with an exponentially decaying amplitude, a so-called "damped sinusoid." Factor ω is seen to be the *natural frequency* of the oscillation in radians per second. Factor α is the *damping coefficient* (per second) and $A\,e^{-\alpha t}$ defines the envelope of the response. If α is large, the response dies out rapidly. Amplitude A and phase angle θ are constants[†] to be determined from initial conditions.

Example 4

(a) Rewrite Eq. 4-17 in Example 3 as a damped sinusoid and determine the natural response after closing the switch in Fig. 4.9.

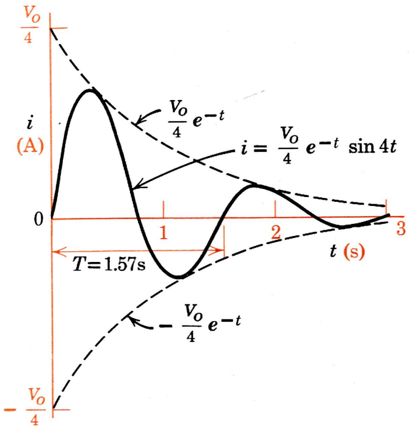


Figure 4.10 Response of an oscillatory second-order system.

For

$$\alpha = \frac{R}{2L} = 1 \quad \text{and} \quad \omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = 4$$

The natural response (Eq. 4-21) is

$$i = A\,e^{-t}\sin(4t + \theta)$$

and

$$\frac{di}{dt} = A\,e^{-t}4\cos(4t + \theta) - A\,e^{-t}\sin(4t + \theta)$$

To evaluate the constants, we recall that just after the switch is closed the current must be zero (since $i_L = 0$) and the rate of change of current must be $V_O/L = V_O$ (since $-v_C = V_O = v_L = L\,di/dt$). At $t = 0^+$,

$$i = 0 = A\,e^0\sin(0 + \theta) = A\sin\theta$$

A is finite; therefore $\theta = 0$ and

$$\frac{di}{dt} = V_O = A\,e^0 4\cos(0) - A\,e^0\sin(0) = 4A$$

Therefore, $A = V_O/4$ and

$$i = \frac{V_O}{4}e^{-t}\sin 4t \tag{4-22}$$

(b) Determine the period of the oscillation and the damping coefficient.

The period (see Fig. 4.10) is

$$T = \frac{1}{f} = \frac{2\pi}{\omega} = \frac{2\pi}{4} = 1.57\text{ s}$$

The damping coefficient is $\alpha = 1\text{ s}^{-1}$; in one second the envelope is down to $1/e$ times its initial value.

Roots Real and Equal. A system described by an equation with complex roots[‡] is said to be "oscillatory" or "underdamped"; in contrast, when the roots are real the system is said to be "overdamped." The limiting condition for oscillation, called the "critically damped" case, occurs when the roots are real and equal. For this case, the discriminant is equal to zero. Physically this is not an important case since it merely represents the borderline between the two regimes.

[†] To obtain Eq. 4-21 from Eq. 4-20, let $B_1 = A\sin\theta$ and $B_2 = A\cos\theta$.

[‡] Since for this case ω^2 is a negative number, the imaginary terms of the roots will be equal and opposite or the roots are always *complex conjugates;* in other words, complex roots always appear in conjugate pairs.

Mathematically it is an interesting case, requiring a special form of solution. If the roots are equal, $s_2 = s_1 = s$. Following the general procedure,

$$i = A_1 e^{st} + A_2 e^{st} = (A_1 + A_2) e^{st} = A e^{st}$$

but this is not sufficiently general for a second-order differential equation. For this case a second term must be included[†] and the general solution is

$$i = A_1 e^{st} + A_2 t e^{st} \tag{4-23}$$

For the RLC circuit of Example 4 with $L = 1$ H and $C = 0.25$ F, the discriminant is zero for

$$R^2 = \frac{4L^2}{LC} = \frac{4}{0.25} = 16 \quad \text{or} \quad R = \sqrt{16} = 4\,\Omega$$

Plotting this response is left for the student. (See Exercise 17.)

Roots in the Complex Plane. We have observed that the character of the natural response of a second-order system is determined by the roots of the characteristic equation. Where the roots are real, negative, and distinct, the response is the sum of two decaying exponentials and the response is said to be overdamped. Where the roots are complex conjugates, the natural response is an exponentially decaying sinusoid and the response is said to be underdamped or oscillatory.

For a general view of this behavior that will give added insight into the response of higher-order systems, consider the effect on the location of the roots of a change in one system parameter. In general, roots are located in the complex plane, the location being defined by coordinates measured along the real or σ (sigma)[‡] axis and the imaginary or $j\omega$ axis. This is referred to as the s plane or, since s has the units of frequency, as the *complex frequency plane.* Specifically, let us consider the locus of the roots for a series RLC circuit as R varies from zero to infinity.

For this circuit the characteristic equation (Eq. 4-13) is

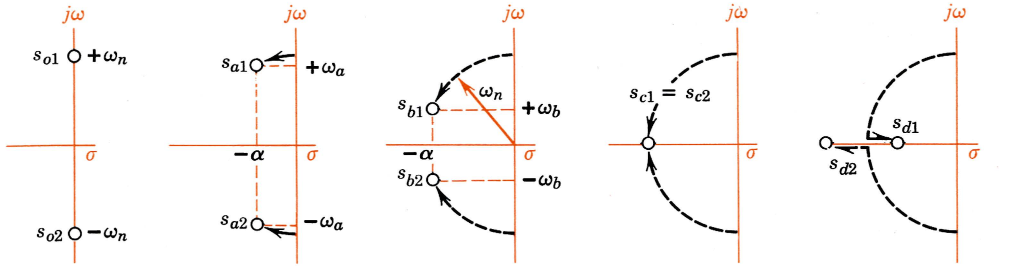
$$Ls^2 + Rs + \frac{1}{C} = 0$$

and the roots are

$$s = -\frac{R}{2L} \pm j\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = -\alpha \pm j\omega$$

[†] See H. B. Phillips, *Differential Equations,* 3rd ed., John Wiley & Sons, New York, 1951, or any other differential equations text.

[‡] The damping coefficient α was defined as $R/2L$, a positive number, so that the decaying nature of $e^{-\alpha t}$ would be evident; positive values of α correspond to negative values of σ and lie in the left half of the complex plane.



$R = 0$
roots imaginary
(o)

$R = R_a$
roots complex
(a)

$R = R_b > R_a$
roots complex
(b)

$R = R_{\text{critical}}$
roots identical
(c)

$R = R_d > R_c$
roots unequal
(d)

Figure 4.11 The root locus for an *RLC* circuit with *R* as the parameter.

For $R = 0$, $s = \pm j\sqrt{1/LC} = \pm j\omega_n$, defining values of the *undamped natural frequency* ω_n. In other words, for $R = 0$, $\alpha = 0$ and the response is oscillatory with no damping. The roots corresponding to the undamped case are located on the imaginary axis of Fig. 4.11o. For $R = R_a$, small but finite, a real part of s appears and the values of s are complex conjugates. As R increases to R_b, the roots move along a path to points s_b. The radial distance from the origin to either root is given by

$$\sqrt{\alpha^2 + \omega^2} = \sqrt{\left(-\frac{R}{2L}\right)^2 + \left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)} = \sqrt{\frac{1}{LC}} = \omega_n \qquad (4\text{-}24)$$

a constant. Therefore, the locus is a circular arc of radius ω_n.

The critical value of resistance is defined by $\omega = 0$ or, in other words, where the discriminant equals zero. For this condition $\sigma = -R/2L = -\sqrt{1/LC}$. At this value of $R = R_c$, the roots coincide on the real axis. As R increases to R_d, s_{d_1} moves toward the origin and s_{d_2} moves out along the negative real axis. As R increases without limit, s_1 approaches the origin and s_2 increases without limit.

With a different circuit parameter the locus takes on a different shape (see Exercise 22). In any such plot, called a *root locus,* the location of the roots determines the character of the natural response, and the effect of changes in a parameter is clearly visible to the initiated. The first step in developing this powerful tool is mastery of a new concept—*impedance.*

Impedance Concepts

If we limit the discussion to exponential waveforms, some interesting and valuable relations between voltage and current can be established. Actually this is not a severe restriction; already we have used exponentials to represent sinusoids, and exponentials with $s = 0$ can be used to represent direct currents. The range of exponential functions for positive and negative, large and small, real values of s is indicated in Fig. 4.12.

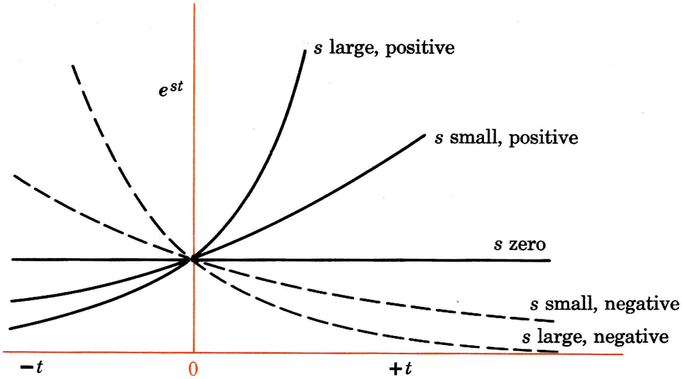


Figure 4.12 The range of exponential functions.

The key property of an exponential function is that its time derivative is also an exponential. For example, if

$$i = I_0 e^{st} \qquad \frac{di}{dt} = sI_0 e^{st} = si \tag{4-25}$$

or if

$$v = V_0 e^{st} \qquad \frac{dv}{dt} = sV_0 e^{st} = sv \tag{4-26}$$

This property greatly facilitates calculating the response of circuits containing resistance, inductance, and capacitance because of the simple voltage-current relations that result.

Impedance to Exponentials

The ratio of voltage to current for exponential waveforms is defined as the *impedance* Z. For a resistance, $v = Ri$ and

$$Z_R = \frac{v}{i} = \frac{Ri}{i} = R \text{ in ohms} \tag{4-27}$$

For an inductance, $v = L(di/dt) = sLi$ and

$$Z_L = \frac{v}{i} = \frac{sLi}{i} = sL \text{ in ohms} \tag{4-28}$$

For a capacitance, $i = C(dv/dt) = sCv$ and

$$Z_C = \frac{v}{i} = \frac{v}{sCv} = \frac{1}{sC} \text{ in ohms} \tag{4-29}$$

The impedances Z_R, Z_L, and Z_C are constants of proportionality between voltages and currents that are exponential functions of time t and frequency s. It can be demonstrated that in each case the dimensions of impedance are the same as those of resistance. (Can you do it?) The general relation

$$v = Zi \tag{4-30}$$

corresponds to Ohm's law for purely resistive circuits, but it must be emphasized that impedance is defined *only for exponentials* and for waveforms that can be represented by exponentials.[†] With this restriction, impedances can be combined in series and parallel just as resistances are, and network theorems can be extended to include circuits containing L and C.

Example 5

In the circuit of Fig. 4.13, $R_1 = 2\,\Omega$, $C = 0.25$ F, $R = 4\,\Omega$. Using the impedance concept, find the currents i and i_R for a voltage $v = 6\,e^{-2t}$ V.

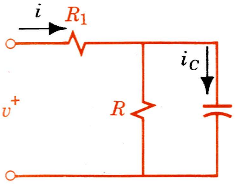


Figure 4.13 The impedance concept in circuit analysis.

Following the rule for resistive networks,

$$Z = Z_1 + \frac{Z_R Z_C}{Z_R + Z_C} = R_1 + \frac{R(1/sC)}{R + (1/sC)}$$

$$= 2 + \frac{4/(-2 \times 0.25)}{4 + 1/(-2 \times 0.25)} = 2 + \frac{-8}{2} = -2\,\Omega$$

Then

$$i = \frac{v}{Z} = \frac{6\,e^{-2t}}{-2} = -\,3\,e^{-2t}$$

Using the current-divider and impedance concepts,

$$i_C = \frac{Z_R \cdot i}{Z_R + Z_C} = \frac{4 \cdot i}{4 - 2} = -6\,e^{-2t}$$

The Impedance Function

For the circuit of Fig. 4.14, the governing equation is

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = v \tag{4-31}$$

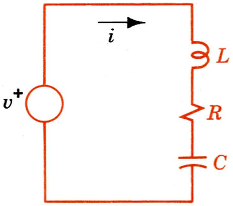


Figure 4.14 The impedance of an RLC circuit.

For an exponential current $i = I_0\,e^{st}$, this becomes

$$Lsi + Ri + \frac{1}{sC}i = v$$

and

$$Z = \frac{v}{i} = sL + R + \frac{1}{sC} \tag{4-32}$$

Note that when the characteristic equation (Eq. 4-13) is divided through by s,

---

[†] Assume another waveform such as $i = I_0 t$; then $v_L = L(di/dt) = LI_0$ and the ratio $v_L/i = LI_0/I_0 t = L/t$ a function of time and *not* a constant as it is for exponentials.

Eq. 4-32 appears; but Eq. 4-32 was obtained by expressing the impedance as a ratio of an exponential voltage to an exponential current. The same relation could also have been obtained by considering that the resultant of impedances in series is the algebraic sum of the impedances or

$$Z = Z_L + Z_R + Z_C = Z(s)$$

where $Z(s)$ signifies the impedance to exponentials of the form $A\,e^{st}$.

Because the impedance function $Z(s)$ contains the same information as the characteristic equation, it is a useful concept in predicting the natural behavior of a system and it can be extended to include the prediction of steady-state response. As indicated in Example 6, the impedance function can be obtained easily in a circuit for which the governing equation may be quite complicated (i.e., may consist of a set of simultaneous integro-differential equations).

Example 6

Given the circuit of Example 5 with $v = V_0\,e^{st}$. Determine and plot $Z(s)$ for real values of s.

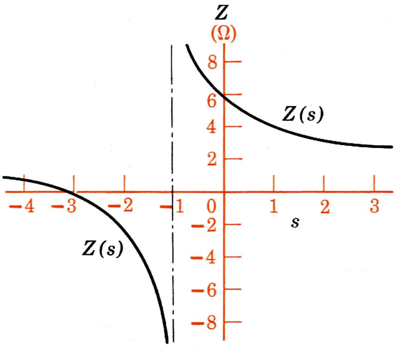


Figure 4.15 The impedance function for the circuit of Figure 4.13.

A great advantage of the impedance concept is the ease with which impedances may be combined. Here

$$Z(s) = Z_1 + \frac{Z_R Z_C}{Z_R + Z_C} = R_1 + \frac{R(1/sC)}{R + 1/sC}$$

$$= R_1 + \frac{R}{RsC + 1} = \frac{sR_1RC + R_1 + R}{sRC + 1}$$

For $R_1 = 2\Omega$, $R = 4\Omega$, and $C = 0.25\text{F}$,

$$Z(s) = \frac{2s + 6}{s + 1} = 2\frac{s + 3}{s + 1}$$

From the graph of Fig. 4.15 and from the expression for $Z(s)$ it is seen that $Z(s) = 0$ at $s = -3$, and that $Z(s)$ increases without limit as s approaches -1.

Poles and Zeros

In Example 6, $Z(s)$ is plotted for real values of s. Since each value of s corresponds to a particular exponential function, the impedance $Z(s_1)$ is the ratio of voltage to current for the particular exponential function of time $i = I_0\,e^{s_1 t}$. For instance, to determine the opposition to direct current flow, the impedance is calculated for $s = 0$ since $i = I_0\,e^{(0)t} = I_0$, a direct current. In Example 6, $Z(0) = 6\,\Omega$, the equivalent of the two resistances in series. This is physically correct because under steady conditions with no change in the voltage across the capacitor, no current flows in C and the circuit, in effect, consists only of the two resistors.

Every point in the complex frequency plane defines an exponential function and the complete plane represents all such functions. The magnitude of the impedance, $|Z(s)|$, can be plotted as a vertical distance above the s plane and, in general, the result will be a complicated surface.

Example 7

Plot the magnitude of the impedance function of the circuit of Example 5 for real values of s and for imaginary values of s, and then sketch the impedance surface.

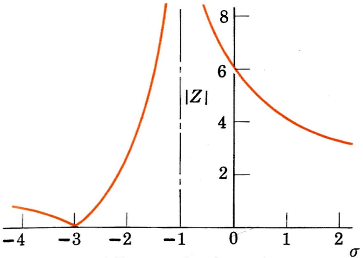
The magnitude of $Z(s)$ for real values of s corresponds to the graph of Fig. 4.15 with negative quantities plotted above the axis. For imaginary values of $s(\pm j1, \pm j2$, etc.) the magnitude of $Z(s)$ is calculated and plotted in a similar way. For $s = -j$,

$$Z(s) = 2\,\frac{3-j}{1-j} \quad \text{and} \quad |Z| = \frac{2\sqrt{10}}{\sqrt{2}} \cong 4.5\,\Omega$$

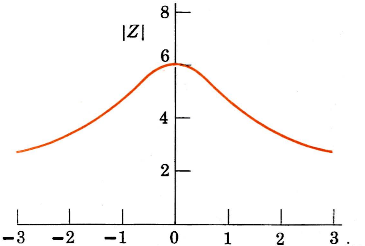
For $s = -1 - j2$,

$$Z(s) = 2\,\frac{2-2j}{0-j2} \quad \text{and} \quad |Z| = \frac{2\sqrt{8}}{\sqrt{4}} \cong 2.8\,\Omega$$

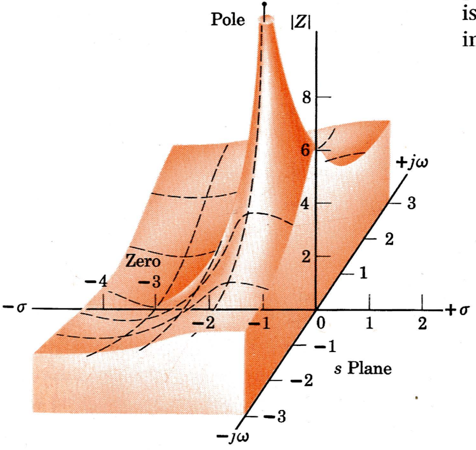
Using the profiles of Figs. 4.16a and b and additional points in the complex plane, the impedance surface is sketched in Fig. 4.16c. The points of zero and infinite impedance are plotted in Fig. 4.16d.



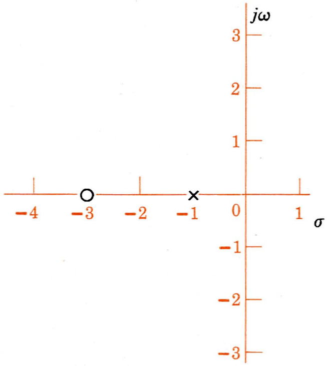
(a) |Z| for real values of s



(b) |Z| for imaginary values of s



(c) |Z| as a three-dimensional surface



(d) Pole–zero diagram

Figure 4.16 Graphical representation of the impedance function.

Typically, the impedance function in three dimensions has the appearance of a tent pitched on the s plane. The height of the tent (magnitude of Z) becomes very great (approaches infinity) for particular values of s appropriately called *poles*. The tent touches the ground at particular values of s called *zeros*. The equation of the surface is usually quite complicated, but the practical use of the impedance function in network analysis or synthesis is relatively simple for two reasons. In the first place, we are usually interested in only a single profile of the surface; for example, the profile along the $j\omega$ axis indicates the response to sinusoidal functions of various frequencies. Second, just as two points define a straight line and three points define a circle, the poles and zeros uniquely define the impedance function, except for a constant scale factor. As a result, the *pole-zero diagram* of Fig. 4.16d contains the essential information of the impedance function shown in Fig. 4.16c. The locations of the poles (**X**) and zeros (**O**), which are relatively easy to find, tell us much about the natural response.

Physical Interpretation of Poles and Zeros

In the general relation between exponential voltage and current

$$v = Zi$$

what is the meaning of "zero impedance"? If the impedance is small, a given current can exist with a small voltage applied. Carrying this idea to the limit, under the conditions of zero impedance† a current can exist with *no applied voltage;* but a current with no applied voltage is, by definition, a natural current response. We conclude, therefore, that each zero, $s = s_1$, of the impedance function for any circuit designates a possible component, $I_1 e^{s_1 t}$, of the natural response current of that circuit. This conclusion is supported by the fact that for the circuit of Fig. 4.14, the zeros of the impedance function (Eq. 4-32) are identical with the roots of the characteristic equation (Eq. 4-13). Knowing the locations of the impedance zeros, we can immediately identify the exponential components of the natural current. (See Example 8.)

If an impedance zero indicates the possibility of a current without a voltage, what is the significance of an impedance pole? Since $v = Zi$, if the impedance is very large, a given voltage can exist with only a small current flowing. Carrying this idea to the limit, under conditions of infinite impedance a voltage can exist

---

†To get the "feel" of zero impedance in a physical system, perform the following mental experiment. Grasp the bob of an imaginary pendulum between your thumb and forefinger. Applying a force, cause the bob to move along its arc with various motions. For motions against gravity or against inertia, appreciable force is required; the mechanical impedance is appreciable. To maintain a purely sinusoidal motion at the natural frequency of the pendulum only a little force (just that needed to make up friction losses) is required; the mechanical impedance to such a motion is small. To maintain a motion approximating an exponentially decaying sinusoid of just the right frequency, *no* force is required; for this motion, velocity is possible with no applied force. In other words, to this motion mechanical impedance is *zero*!

Example 8

Given the circuit shown in Fig. 4.17, determine the poles and zeros of impedance. If energy is stored in the circuit in the form of an initial voltage V_O on the capacitor, predict the current i that will flow when the switch S is closed.

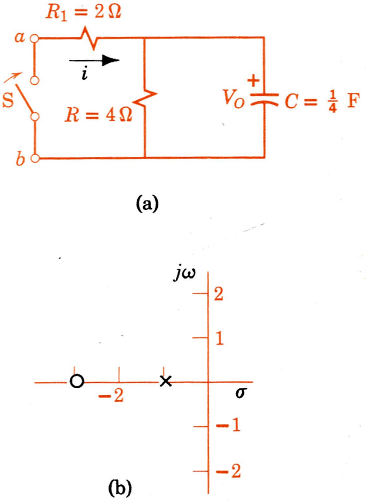


Figure 4.17 A pole-zero diagram of impedance.

As previously determined, the impedance function looking into the circuit at terminals ab is

$$Z(s) = 2\frac{s+3}{s+1}$$

$$\begin{aligned} \text{Where } s &= -1, \text{ the denominator is zero and} \\ Z(s) &= \infty; \text{ therefore,} \\ s &= -1 \text{ is a pole.} \\ \text{Where } s &= -3, \text{ the numerator is zero and} \\ Z(s) &= 0, \text{ therefore,} \\ s &= -3 \text{ is a zero.} \end{aligned}$$

If the impedance is zero, a current can exist with no external forcing voltage. Therefore, the natural current behavior is defined by $s = s_1 = -3$ or

$$i = I_1 e^{-3t}$$

As before, I_1 is evaluated from initial data. At the instant the switch is closed, V_O appears across R_1 (tending to cause a current opposite to that assumed) and

$$i_O = I_1 e^0 = I_1 = -\frac{V_O}{R_1}$$

Hence,

$$i = -\frac{V_O}{R_1} e^{-3t}$$

is the natural response current.

with *no current flow,* in other words, a natural voltage. We conclude that each pole, $s = s_a$, of the impedance function of a circuit designates a possible component, $V_a e^{s_a t}$, of the natural voltage response of that circuit. (See Example 9 on page 106.)

The General Impedance Function

Typical impedance functions include (for the circuit of Example 6)

$$Z(s) = R_1 + \frac{R(1/sC)}{R + (1/sC)} = \frac{sR_1RC + R_1 + R}{sRC + 1}$$

and (for the circuit of Figure 4.14)

$$Z(s) = sL + R + \frac{1}{sC} = \frac{s^2LC + sRC + 1}{sC}$$

The impedance function for any network, no matter how complicated, consisting of resistances, inductances, and capacitances, can be reduced to the ratio of two

Example 9

Returning to Example 8, assume that energy is stored in the form of an initial voltage V_0 on the capacitor (perhaps by means of the voltage source shown in Fig. 4.18). Predict the voltage v that will appear across terminals ab when the switch S is opened.

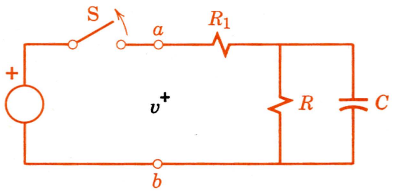


Figure 4.18 Natural voltage response.

With the external energy source removed, only a natural behavior voltage can appear. Such a voltage can exist with no current flow only if the impedance is infinite. For a pole at $s = s_a = -1$, the natural voltage behavior is

$$v = V_a e^{-t}$$

V_a is evaluated from initial data. At the instant the switch is opened, the current in R_1 goes to zero and the voltage across terminals ab is just $v_R = v_C = +V_0$. Hence,

$$v = V_0 e^{-t}$$

is the natural response voltage. This result can be checked by letting $v = i_2 R$ where i_2 is the natural current that would flow in R if R were suddenly connected across the charged capacitor.

polynomials in s. In general, we can write

$$Z(s) = K\frac{s^n + \cdots + k_2 s^2 + k_1 s + k_0}{s^m + \cdots + c_2 s^2 + c_1 s + c_0} \tag{4-33}$$

Although it may not be easy, this can always be factored into

$$Z(s) = K\frac{(s - s_1)(s - s_2)\cdots(s - s_n)}{(s - s_a)(s - s_b)\cdots(s - s_m)} \tag{4-34}$$

When $s = s_1, s_2, \ldots, s_n$, $Z(s) = 0$; therefore, these are zeros.
When $s = s_a, s_b, \ldots, s_m$, $Z(s) = \infty$; therefore, these are poles.

If the network is known, the impedance function can be written and factored and the pole-zero diagram constructed. Conversely, from the pole-zero diagram the impedance function can be formulated (except for scale factor K in Eq. 4-34). Theoretically, a network can then be designed or "synthesized," but practically this is not always easy. The complicated problems are so difficult that entire books have been written on the subject of network synthesis. In this book, however, we are concerned primarily with network analysis and the problems are more straightforward.

The General Admittance Function

Impedance is defined for exponentials as the ratio of voltage to current. The reciprocal of impedance is *admittance*, a useful property defined as the ratio of exponential current in amperes to voltage in volts so that

$$Y = \frac{i}{v} = \frac{1}{Z} \tag{4-35}$$

measured in mhos (℧, reciprocal of ohms). For exponential voltages and currents, the admittances of ideal circuit elements are:

$$\begin{aligned} Y_R &= \frac{i_R}{v_R} = \frac{i}{Ri} = \frac{1}{R} = G \\ Y_L &= \frac{i_L}{v_L} = \frac{i}{L(di/dt)} = \frac{i}{sLi} = \frac{1}{sL} \\ Y_C &= \frac{i_C}{v_C} = \frac{C(dv/dt)}{v} = \frac{sCv}{v} = sC \end{aligned} \qquad (4\text{-}36)$$

Admittance is particularly useful in analyzing circuits that contain elements connected in parallel. Since current is directly proportional to admittance ($i = Yv$), admittances in parallel can be added directly just as conductances in parallel are added.

The admittance function $Y(s)$ for any network consisting of lumped passive elements also can be reduced to the ratio of two polynomials in s. In the standard factored form this becomes

$$Y(s) = \frac{1}{Z(s)} = \frac{1}{K} \frac{(s - s_a)(s - s_b) \cdots (s - s_m)}{(s - s_1)(s - s_2) \cdots (s - s_n)} \qquad (4\text{-}37)$$

Note that the admittance function has poles where the impedance function has zeros, and vice versa. The pole-zero diagram for the admittance function contains the same information as the pole-zero diagram for the impedance function, but the diagrams are labeled differently.

General Procedure for Using Poles and Zeros

The pole-zero concept is a powerful tool in determining the natural behavior (and, as we shall see in Chapter 5, the forced behavior) of any linear physical system. Modified to take advantage of this concept, the general procedure for determining the natural behavior of an electrical circuit is:

1. Write the impedance or admittance function for the terminals of interest.
2. Determine the poles and zeros, and plot the pole-zero diagram.
3. (a) For the terminals short-circuited, the natural behavior current is

   $$i = I_1 e^{s_1 t} + I_2 e^{s_2 t} + \cdots + I_n e^{s_n t} \qquad (4\text{-}38)$$

   where $s_1, s_2, \ldots, s_n$ are zeros of the impedance function or poles of the admittance function.

   (b) For the terminals open-circuited, the natural behavior voltage is

   $$v = V_a e^{s_a t} + V_b e^{s_b t} + \cdots + V_m e^{s_m t} \qquad (4\text{-}39)$$

   where $s_a, s_b, \ldots, s_m$ are poles of the impedance function or zeros of the admittance function.
4. Evaluate the coefficients from the initial conditions (Example 10, p. 108).

Example 10

(a) In the circuit of Fig. 4.19a, the voltage source has been connected for a time long enough for steady-state conditions to be reached. At time $t = 0$, switch S is opened. Predict the open-circuit voltage across terminals *ab*.

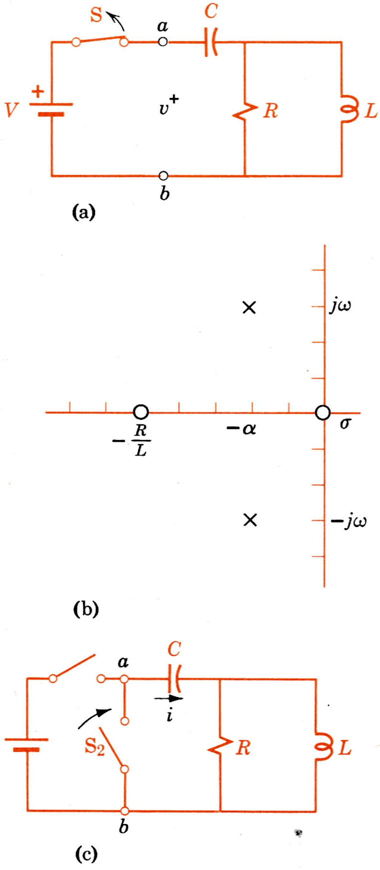


Figure 4.19 The use of poles and zeros.

Following the general procedure,

1. $$Z_{ab}(s) = \frac{1}{sC} + \frac{RsL}{R + sL} = \frac{R + sL + s^2RLC}{sC(R + sL)}$$

$$= R\,\frac{s^2 + (1/RC)s + 1/LC}{s(s + R/L)} = \frac{1}{Y(s)}$$

In the general form of Eq. 4.37,

$$Y(s) = \frac{1}{R}\,\frac{(s - 0)(s - [-R/L])}{(s - s_1)(s - s_2)}$$

2. The admittance function has zeros at $s = 0$ and $s = -R/L$, and poles at $s = s_1$ and $s = s_2$ where

$$s_1, s_2 = -\frac{1}{2RC} \pm \sqrt{\frac{1}{4R^2C^2} - \frac{1}{LC}}$$

Assuming complex poles, the pole-zero diagram of admittance is as shown in Fig. 4.19b.

3. For the terminals open-circuited, the natural behavior voltage is defined by the zeros of $Y(s)$ or

$$v = V_a\,e^0 + V_b\,e^{-(R/L)t}$$

4. At $t = 0^-$, the current in the inductance is constant and $v_L = L\,di/dt = 0 = v_R \therefore$ all the voltage V appears across the capacitance and the current in the inductance is zero. The second component of voltage (which reflects the *possibility* of energy storage in the inductance) is zero.

The only energy storage is in the capacitor; at $t = 0^+$,

$$v = V_a\,e^0 = V$$

or the open-circuit voltage across terminals *ab* is just V. (There is no way for this ideal capacitor to discharge.)

(b) Derive an expression for current i if the terminals *ab* are short-circuited by means of switch S_2 as shown in Figure 4.19c.

For terminals shorted, the natural behavior current is defined by the poles of $Y(s)$ or

$$i = I_1\,e^{s_1 t} + I_2\,e^{s_2 t}$$

The current response will be damped or oscillatory depending on the numerical value of the discriminant.

SUMMARY

- Forced behavior is the response to external energy sources.
  Natural behavior is the response to internal stored energy.

- Many physical systems with one energy-storage element can be described adequately by first-order integro-differential equations. The general procedure for determining the natural behavior of a linear system is:

  1. Write the governing integro-differential equation.
  2. Reduce this to a homogeneous differential equation.
  3. Assume an exponential solution with undetermined constants.
  4. Determine the exponents from the homogeneous equation.
  5. Evaluate the coefficients from the given conditions.

- In a second-order system with two energy-storage elements, the character of the natural response is determined by the discriminant.
  If the discriminant is *positive*, the response is *overdamped* and is represented by the sum of two decaying exponentials: $a = A_1 e^{s_1 t} + A_2 e^{s_2 t}$.
  If the discriminant is *negative*, the response is *oscillatory* and is represented by the damped sinusoid: $a = A e^{-\alpha t} \sin(\omega t + \theta)$.
  If the discriminant is *zero*, the response is *critically damped* and is represented by the sum of two different terms: $a = A_1 e^{st} + A_2 t e^{st}$.

- Impedance Z (ohms) and admittance Y (mhos) are defined for exponentials.

  Where $v = Zi$, $Z_R = R$, $Z_L = sL$, and $Z_C = 1/sC$
  Where $i = Yv$, $Y_R = G$, $Y_L = 1/sL$, and $Y_C = sC$

- Impedances and admittances in complicated networks are combined in the same way as resistances and conductances, respectively.
  The impedance function $Z(s)$ and the admittance function $Y(s)$ contain the same information as the characteristic equation.

- The pole-zero diagram contains the essential information of the impedance function or the admittance function.
  A zero of the impedance function indicates the possibility of a current without an applied voltage, therefore a natural current.
  A pole of the impedance function indicates the possibility of a voltage without an applied current, therefore a natural voltage.

- Using the pole-zero concept, the general procedure is:

  1. Write the impedance or admittance function for the terminals of interest.
  2. Determine the poles and zeros of impedance or admittance.
  3. Use the poles and zeros to identify possible components of natural voltage or current.
  4. Evaluate the coefficients from the given conditions.

REVIEW QUESTIONS

1. Cite an example of natural behavior in each of the following branches of engineering: aeronautical, chemical, civil, industrial, and mechanical.
2. To what extent is the natural behavior of a system influenced by the waveform of the forcing function that stores energy in the system?
3. Discuss the possibility of a positive exponent appearing in the natural response of a passive network. In any physical system.
4. Outline the procedure for determining the natural behavior of a mechanical system in translation.
5. In contrast to the exponential behavior of its idealized model, an actual coasting automobile comes to a complete stop long before an infinite time. Why?
6. Measurements on a certain very fast transient are difficult and the results include large random errors. In determining the time constant for this system, would a linear or a semilog plot of experimental values be preferable? Why?
7. What is the physical difference between a "first-order" system and a "second-order" system?
8. In what sense is the "characteristic equation" characteristic of the circuit?
9. In what sense does the "discriminant" discriminate?
10. What important initial information is available from an inductor in a circuit? From a capacitor? Explain.
11. In the case of a practical resistor, can the current be changed instantaneously? Can the voltage? Explain.
12. How do the waveforms represented by $i_1 = I_0 e^{j4t}$ and $i_2 = I_0(e^{j4t} + e^{-j4t})$ differ? Could both be plotted on the same graph? Explain.
13. A damped sinusoid has a period, but is it a "periodic wave"? Explain.
14. Why is the natural response determined by the homogeneous equation instead of by the governing integro-differential equation?
15. What is the meaning of a "complex" frequency?
16. Give three reasons why so much attention is devoted to exponential functions in network analysis?
17. If, for exponentials, si corresponds to di/dt, what is the significance of $(1/s)i$?
18. What is the value of $|Z|$ (height of the "tent") in Fig. 4.16 for large real values of s? For large imaginary values? What is the limit of $|Z|$ as s approaches infinity?
19. In regard to *natural* response, is there any information available in the impedance function (Fig. 4.16c) that is not in the pole-zero diagram (Fig. 4.16d)? Explain.
20. Devise and sketch a physical system in which there is force without velocity (corresponding to an infinite mechanical impedance) and in which the force decays exponentially with time.

EXERCISES

1. In Fig. 4.20, switch S has been closed a long time.
   (a) If conditions have reached the steady state, what is the voltage across L? The current through R? The current through L?
   (b) If switch S is opened at time $t = 0$, write the equation governing current i as a function of time.
   (c) Solve for $i(t)$ and sketch $i(t)$ on a graph.
   (d) Determine $v(t)$ and sketch on the graph of part (c).

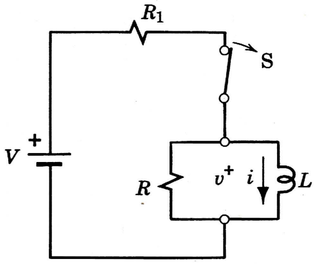


Figure 4.20

2. The energy stored in a 100-μF capacitance is 80 mJ. At time $t = 0$, a 5-kΩ resistance is connected across the capacitance.
   (a) What is the initial voltage on the charged capacitance?
   (b) Draw the circuit showing a switch that is closed at $t = 0$, and predict the voltage across C and the current through R at $t = 0^+$.
   (c) Express the current *out* of C and the current *into* R in terms of v_C and write the governing equation in terms of v_C.
   (d) Assume an exponential solution for $v_C(t)$, show that it satisfies the governing equation and the initial condition, and predict v_C at $t = 1$ s.
   (e) From your expression for $v_C(t)$, derive an expression for $i_R(t)$, evaluate the time constant T, and predict i_R at $t = 1.5$ s.
   (f) Calculate the total energy dissipated in resistance R. Explain the significance of this value.
3. In Fig. 4.21, C_2 is uncharged and switch S has been in position 1 a long time.
   (a) Under steady-state conditions, what is the current in C_1? The current in R_1? The voltage across C_1?
   (b) If switch S is thrown to position 2 at time $t = 0$, write the equation governing current i as a function of time.
   (c) Assuming $C_1 = C_2$, solve for $i(t)$ and sketch $i(t)$ on a graph.
   (d) Determine $v_{C_2}(t)$ and sketch on the graph of part (c).

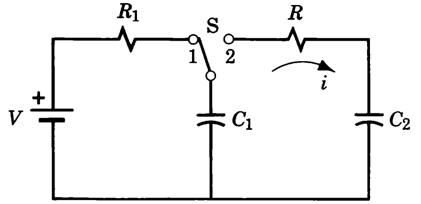


Figure 4.21

4. A flywheel of radius r and moment of inertia J kg-m^2 is being driven at angular velocity ω rad/sec against a friction torque $T_f = D\omega$ N-m. At time $t = 0$, the driving torque T_d is removed and the flywheel coasts.
   (a) Write the differential equation governing the coasting period.
   (b) Derive an equation for ω as a function of time.
5. Two bicycle wheels are similar in construction, but one has a time constant twice that of the other. Which will be easier to pedal? Why?
6. In Fig. 4.22, switch S has been closed a long time. At time $t = 0$, the switch is opened.
   (a) Just before $t = 0$, what is the voltage across L? The current through R_2?
   (b) At $t = 0^+$, what is the current i?
   (c) Derive an expression for $i(t)$ for $t > 0$.
   (d) Define the time constant for this circuit.

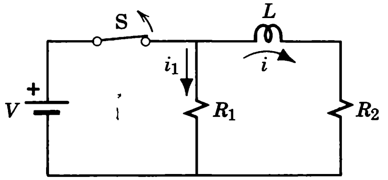


Figure 4.22

7. In Fig. 4.22, $V = 10$ V, $R_1 = 10\,\Omega$, $R_2 = 5\,\Omega$, and $L = 5$ H. Switch S is closed a "long" time and then opened at $t = 0$.
   (a) Determine the values of i_1 and i at $t = 0^+$.
   (b) Sketch approximately to scale i_1 and i as functions of time before and after $t = 0$.
8. After switch S in Fig. 4.21 is thrown to position 2, the source-free circuit consists of two capacitances and a resistance.
   (a) What single capacitance C_{eq} is equivalent to C_1 and C_2 in series?
   (b) What is the time constant of the circuit in which i flows?
   (c) Referring to Eq. 4-11, derive an expression for $i(t)$ in terms of the circuit parameters.
9. In the circuit of Fig. 4.6a, $C_1 = 2\,\mu$F, $C_2 = 3\,\mu$F, $R_1 = 5$ kΩ, $R_2 = 2$ kΩ, and $R_3 = 3$ kΩ.
   (a) Predict the time constant for voltages and currents in this circuit.
   (b) Voltage $V = 21$ V is applied across R_3 for a "long" time and removed at $t = 0$. Define polarities and derive specific equations for $v_{C1}(t)$ and $i_{R2}(t)$.

10. In Fig. 4.23, energy is stored in the circuit in the form of initial currents I_1 and I_2 in the inductances.

(a) Write the governing equations as a pair of simultaneous equations in i_1 and i_2 and their derivatives.

(b) Outline a procedure for determining the natural response currents $i_1(t)$ and $i_2(t)$ from the governing equations.

(c) Would you expect this to be an easy or difficult procedure? (An easier method is suggested in Exercise 31).

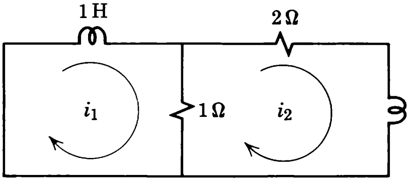


Figure 4.23

11. In Fig. 4.24, switch S_1 has been in position 1 and switch S_2 has been closed for a long time before $t = 0$. The 2-μF capacitor is initially uncharged.

(a) For $t < 0$, find i_L, v_L, i_{C1}, and v_{C1}. (Define voltage polarities consistent with current directions.)

(b) At $t = 0$, switch S_1 is thrown to position 2 and switch S_2 is opened. For $t = 0^+$, find i_L, i_R, i_{C1}, v_{C1}, v_R, v_{C2}, and v_L.

(c) For $t = 0^+$, find di_R/dt, di_L/dt, dv_C/dt, and di_2/dt.

12. In a series RLC circuit, $R = 1250\,\Omega$, $L = 62.5$ H, and $C = 250\,\mu$F. Energy is stored in the form of an initial voltage $V_0 = 25$ V on the capacitor.

(a) Draw the circuit, define polarities, and calculate the numerical values of i, v_R, and v_L at $t = 0^+$.

(b) Write the "governing" equation, the "homogeneous" equation, and the "characteristic" equation for this circuit.

(c) Calculate di/dt and d^2i/dt^2 at $t = 0^+$.

(d) Will the natural response be under, over, or critically damped?

(e) Solve for $i(t)$, evaluating the constants from initial conditions.

13. In the circuit of Exercise 12, at $t = 0$ the capacitance voltage is $V_0 = 20$ V and initial current $I_0 = 40$ mA is flowing in the inductance (tending to discharge the capacitance). Draw the circuit and calculate the numerical values of i, v_C, v_L, v_R, and di/dt at $t = 0^+$.

14. In a series RLC circuit, $R = 666\,\Omega$, $L = 8.33$ H, and $C = 100\,\mu$F. Energy is stored in the form of a 2-A current in the inductor. At $t = 0$, the exciting current is removed and the inductor shorted across R and C (initially uncharged) in series.

(a) Draw the circuit and calculate the numerical values of i, v_C, v_R, and v_L at $t = 0^+$.

(b) Calculate di/dt and d^2i/dt^2 at $t = 0^+$.

(c) For this system, write the "governing" equation, the "homogeneous" equation, and the "characteristic" equation.

(d) Will the natural response be under-, over-, or critically damped?

(e) Solve for $i(t)$, evaluating all constants from initial conditions.

15. Explain the natural response of an oscillatory RLC circuit on a physical basis in terms of the energy interchanges that occur. (You may wish to reason by analogy with an oscillating pendulum.)

16. In Fig. 4.25, switch S has been in position 1 for a "long" time. At time $t = 0$, the switch is moved to position 2.

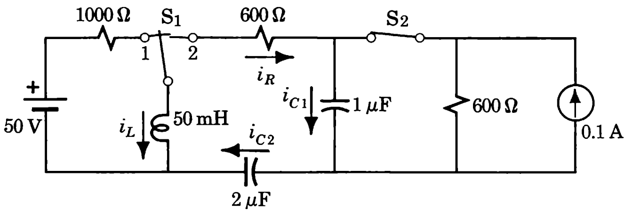


Figure 4.24

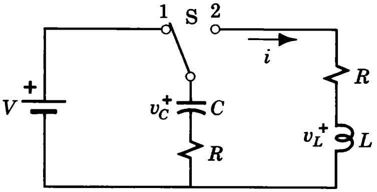


Figure 4.25

(a) For time $t = 0^+$, evaluate v_C, i, and v_L.
(b) Derive an expression for current i as a function of time $t > 0$ in terms of the *given* quantities.

17. An *RLC* series circuit has an initial current I_0 in the inductor and no voltage on the capacitor.
(a) For the case of critical damping, derive a general expression for the natural current response.
(b) For $I_0 = 2$ A, $L = 1$ H, $C = 0.25$ F, and $R = 4\ \Omega$, determine and plot to scale the response current as a function of time.

18. In the circuit of Fig. 4.26, $V = 10$ V, $L = 1$ H, $R = 5\ \Omega$, and $C = 0.1$ F. The switch is thrown to position 2 at $t = 0$.

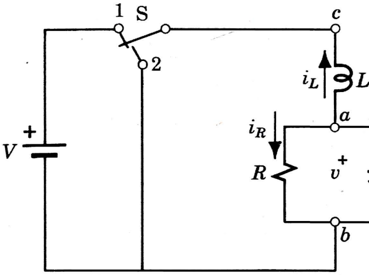


Figure 4.26

(a) Assuming the switch was in position 1 a "long" time, evaluate i_L, i_R, and $v_C = v$ just *before* the switch is thrown and just *after* the switch is thrown. Where is energy stored in the system?
(b) By summing the currents out of node *a*, write the equation governing the behavior of this circuit for $t > 0$ in terms of $v = v_{ab}$. Use this equation and information from part (a) to evaluate dv/dt at $t = 0^+$. (Hint: What is the value of $\int i\,dt$ at $t = 0^+$?)
(c) From the "governing" equation, write the "homogeneous" equation and the "characteristic" equation.
(d) Use the roots of the characteristic equation to write the voltage response $v(t)$ with undetermined constants, but in a form that has physical significance.
(e) Evaluate the coefficients by writing simultaneous equations using information about v and dv/dt at $t = 0^+$.
(f) Sketch the envelope of the equation, note the intercepts on the time axis, and sketch in $v(t)$ approximately to scale.

19. In the circuit of Fig. 4.27, $L = 0.5$ H, $R = 400\ \Omega$, and $C = 0.5\ \mu$F. The capacitor has an initial voltage $V_0 = 30$ V, and the switch is closed at $t = 0$.
(a) Determine the numerical values of v, i_R, i_L, and i_C just after the switch is closed, i.e., at $t = 0^+$.
(b) Calculate dv/dt and d^2v/dt^2 at $t = 0^+$.
(c) For this system, write the "governing" equation in terms of v, the "homogeneous" equation, and the "characteristic" equation.
(d) Will the response be under-, over-, or critically damped?
(e) Solve for $v(t)$ and evaluate all constants from initial conditions.
(f) Plot enough points to sketch, approximately to scale, the voltage as a function of time. (You may prefer to do this in terms of the components.)

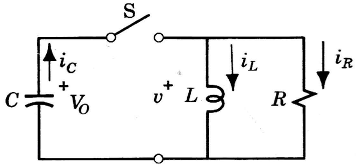


Figure 4.27

20. Repeat Exercise 19 with $L = 64$ mH, $R = 400\ \Omega$, and $C = 0.5\ \mu$F. The switch is closed at $t = 0$ with $V_0 = 30$ V.

21. Draw a "map of *s*-land" similar to but larger than Fig. 4.28. In each square sketch the

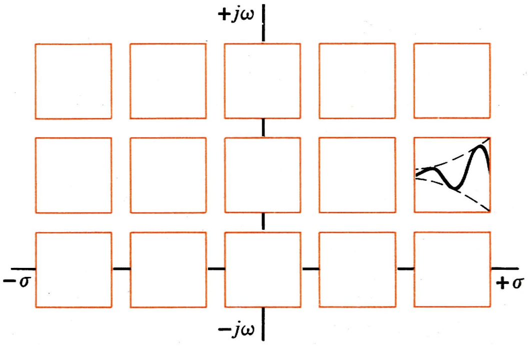


Figure 4.28

response expected for a root of the characteristic equation located at that point in the complex plane. Follow the example drawn for a root with positive values of σ and ω, showing an exponentially increasing sinusoidal function of time, $e^{(\sigma+j\omega)t}$. Clearly indicate differences in the expected response corresponding to differences in σ and ω.

22. Determine and sketch the root locus for a series *RLC* circuit in which:
  (a) C is the changing parameter.
  (b) L is the changing parameter.

23. In the circuit of Fig. 4.29, $R_1 = 2\ \Omega$, $R = 4\ \Omega$, and $L = 2$ H.
  (a) Determine enough points to plot $Z(s)$.
  (b) What is the impedance of this circuit to direct current?
  (c) If a voltage $v = V_0 e^{st}$ is acting where $V_0 = 1$ V, what current i flows for $s = -4$? For $s = -3$? For $s = -2$?
  (d) Assume current $i = 2\,e^{-4t}$ A is flowing and calculate the necessary voltage $v(t)$.

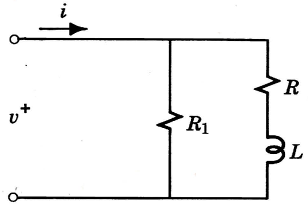


Figure 4.29

24. Given the circuit of Fig. 4.30,
  (a) Define in words the impedance at the terminals.
  (b) Determine the impedance $Z(s)$ at the terminals.
  (c) For $i = 5\,e^{-3t}$ A, evaluate $Z(s)$ and determine v_C the voltage across the capacitance by using the impedance concept.

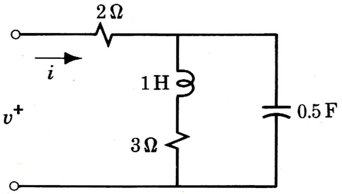
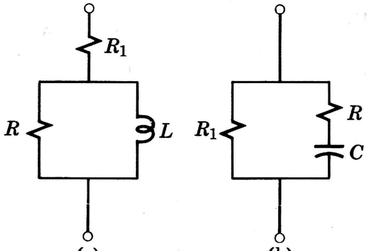


Figure 4.30

  (d) For $i = 5$ A dc, repeat part (c) and check your answer by reasoning from basic principles of L and C.

25. For the circuit of Fig. 4.30,
  (a) Write the impedance function and sketch the pole-zero diagram.
  (b) Write the admittance function and sketch the pole-zero diagram.

26. For the circuits in Figs. 4.31a and b,
  (a) Write the impedance function and sketch the pole-zero diagram.
  (b) Write the admittance function and sketch the pole-zero diagram.



(a) (b)

Figure 4.31

27. For a certain network, the pole-zero diagram of impedance is as shown in Fig. 4.32. Energy is stored in the network and then at time $t = 0$ a switch is thrown to disconnect all external energy sources.
  (a) Write expressions (with undetermined constants) for the short-circuit natural response current and the open-circuit natural response voltage.
  (b) Write an expression for $Z(s)$ for this network.

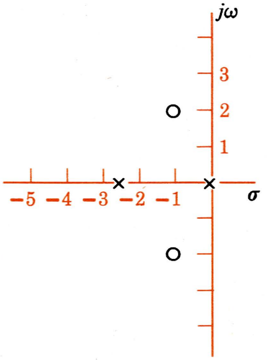


Figure 4.32

(c) Synthesize the network, using one element each of R, L, and C. (Hint: Let $C = 0.01$ F and determine the necessary values of R and L.)

28. For the circuit shown in Fig. 4.33,
(a) Find the impedance function $Z(s)$.
(b) Draw the pole-zero diagram of impedance.
(c) Explain briefly the physical significance of each zero and each pole.

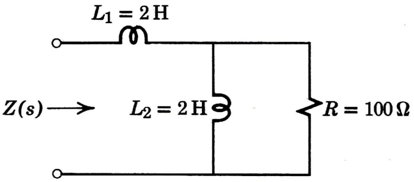


Figure 4.33

29. For the circuit of Exercise 18 (Fig. 4.26) with the switch in position 2,
(a) Evaluate the admittance of the circuit as measured at terminals *ab* in the form of Eq. 4-37. (It may help to redraw the circuit.)
(b) Determine the poles and zeros of admittance and plot a pole-zero diagram.
(c) Write the natural voltage response across these terminals with undetermined constants.
(d) Compare the result of part (c) with that for Exercise 18(d).

30. In the circuit of Fig. 4.34, $V_O = 10$ V, $C = 0.5$ F, $R_1 = R_2 = 2\ \Omega$, and $L = 1$ H.
(a) Outline a step-by-step procedure for obtaining the natural response current i upon closing switch S.
(b) Derive an expression for $i(t)$ with two undetermined constants.

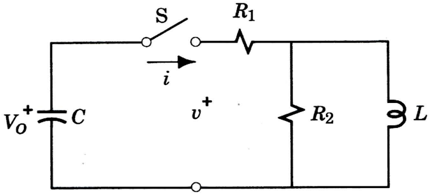


Figure 4.34

31. The necessity for solving simultaneous equations in Exercise 10 is avoided if the impedance concept is employed. Open up the circuit to the left of the 1-H inductance and designate a pair of terminals at which i_1 would be a short-circuit current.
(a) Write the impedance function $Z(s)$, determine the poles and zeros, and plot the pole-zero diagram.
(b) For initial currents $I_1 = I_2 = 3$ A, determine $i_1(t)$ and sketch it approximately to scale.

32. Use the pole-zero concept to find $v(t)$ in the circuit of Exercise 20.
(a) Write the admittance function for the terminals across which voltage v appears. (Hint: Redraw the circuit.)
(b) Determine the poles and zeros of $Y(s)$ and plot.
(c) Write an expression for $v(t)$ with two undetermined constants.
(d) What is the natural frequency of this circuit in Hz?

PROBLEMS

1. In Fig. 4.35 switch S is opened at $t = 0$ after being closed a long time. Predict current i as a function of time after $t = 0$. Sketch $i(t)$ just before $t = 0$ and for a significant time after.
2. A vacuum-insulated bottle contains W lb of a liquid of specific heat C_p (Btu/lb-°F) at a temperature τ_0 (°F above the outside temperature). The effective thermal conductance of the container is U (Btu/s-°F temperature difference).

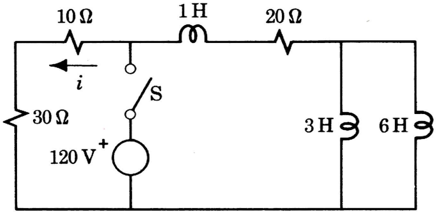


Figure 4.35

(a) Stating all assumptions and checking your equations dimensionally, express the heat stored in the liquid and the rate of heat conduction through the container.

(b) Predict the temperature of the liquid as a function of time.

3. An automobile weighing 3220 lb is moving at a velocity of 50 ft/s. At time $t = 0$, the car is shifted into neutral and allowed to coast. In 2.5 s, the velocity drops to 45 ft/s. Predict the distance the auto will coast on a level road.

4. In an RC circuit undergoing natural behavior, where is energy stored? Where is energy dissipated? Using the expression for current, show that the total energy dissipated during natural behavior is just equal to the energy stored initially.

5. Energy is to be stored in an inductor and dissipated in a 20-Ω resistor. The power in the resistor must be cut in half every second.

(a) Design the inductor, i.e., specify L of an ideal inductance.

(b) If a practical inductor introduces 1 Ω resistance for every 10 H of inductance, specify L.

6. In the circuit of Fig. 4.29, $R_1 = 4\,\Omega$, $R = 2\,\Omega$, and $L = 3$ H.

(a) Determine the impedance of this circuit to an exponential where $s = -2 + j2$.

(b) A voltage $v = 10 \cos (4t - \pi/6)$ V appears across the terminals. Express this voltage as the real part of an exponential and derive an expression for the current i as a function of time.

7. A certain application requires an electrical circuit with the pole-zero diagram of admittance shown in Fig. 4.36.

(a) Create a possible circuit with these general characteristics.

(b) Design the circuit by specifying the component values.

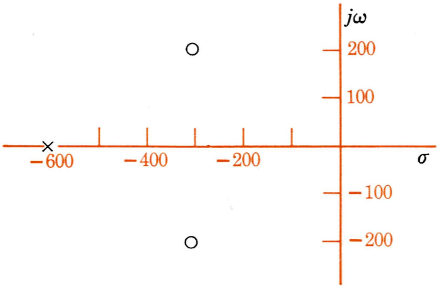


Figure 4.36

8. In the circuit of Fig. 4.34, $C = 1\,\mu$F, $R_1 = R_2 = 2$ kΩ, and $L = 5$ mH. The capacitor carries an initial voltage $V_o = 12$ V. Outline a step-by-step procedure and determine $v(t)$ after the switch is closed.

9. The circuit of Fig. 4.37 is useful in the design of control systems. In this case, $C = 1\,\mu$F, $R = 1$ MΩ, and the initial voltage (on the right-hand capacitor only) is $V_o = 20$ V. Determine the current $i(t)$ after switch S is closed.

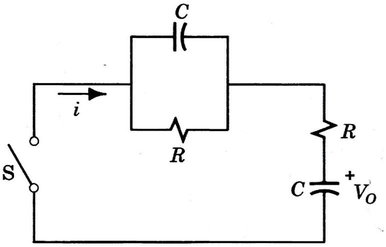


Figure 4.37

10. In Fig. 4.23, energy is stored in the circuit in the form of initial currents $I_1 = 10$ A and $I_2 = 2$ A in the inductances. Determine the current in the 1-Ω resistance as a function of time.

Chapter 5
Forced Response

Forced Responses
Phasor Methods
AC Circuit Analysis
Analogs and Duals

Internal energy storage gives rise to natural behavior; regardless of how the energy is stored or where, the natural response of a circuit is determined by the characteristics of the circuit itself. In contrast, the form of the forced response due to an external energy source is dependent on the form of the forcing function; for example, a sinusoidal voltage always produces sinusoidal currents. Having mastered the technique of predicting natural behavior, we are now ready to tackle the problem of determining the response to some important forcing functions or waveforms. Then in Chapter 6 we shall use our knowledge of natural and forced response to determine the complete response of an electrical circuit, employing an approach that is applicable to nonelectrical problems as well.

In limiting consideration to just the forced response, we are assuming that natural behavior is absent. Either there was no natural response, or sufficient time has elapsed since the initiation of any transient effect for the natural response to become negligibly small. This condition exists in many practical problems.

The material in this chapter represents another step in a cumulative process. We use the accepted definitions, the fundamental element and connection laws, the methods of representing signal waveforms, and the impedance concept to solve a new set of problems. Although impedance was defined for exponentials, the general interpretation permits its application to continuous and sinusoidal waves as well. To simplify the analysis of circuits involving sinusoidal voltages and currents, we use phasors.

Response to Forcing Functions

By definition, impedance is the ratio of voltage to current for exponentials of the form $i = I_0 e^{st}$. In general,

$$v = Z(s)i \tag{5-1}$$

a special form of Ohm's law where impedance $Z(s)$ is in ohms.

For resistance, $v_R = Ri$ and $Z_R(s) = v/i = R.$
For inductance, $v_L = L\,di/dt = sLi$ and $Z_L(s) = v/i = sL.$
For capacitance, $i_C = C\,dv/dt = sCv$ and $Z_C(s) = v/i = 1/sC.$

Impedances to exponentials can be combined in series and parallel just as resistances are.

Response to Exponentials

In the circuit of Fig. 5.1a, an exponential forcing function $i = I_0 e^{-\alpha t}$ flows in the series combination of R and L. There is no information on how the current was initiated; we do know that such a current could not be established by closing a switch at $t = 0$. (Why not?†) Assuming such a current does exist, what is the response of the circuit in terms of voltages across the elements?

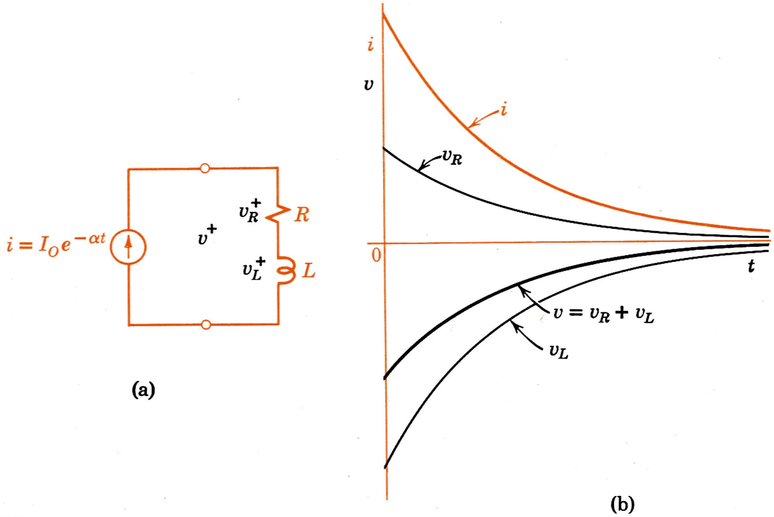


Figure 5.1 Forced response of an *RL* circuit to an exponential current.

Since $Z_R(s) = R$,

$$v_R = Z_R(s)i = RI_0\,e^{-\alpha t}$$

Since $Z_L(s) = sL = -\alpha L$,

$$v_L = Z_L(s)i = -\alpha LI_0\,e^{-\alpha t}$$

---

† See footnote on p. 15.

The total voltage is the sum of the two in series or

$$v = v_R + v_L = (R - \alpha L)I_0\, e^{-\alpha t}$$

As expected, the response to an exponential forcing function is itself exponential of the same "frequency" α or the same time constant $T = 1/\alpha$. The negative sign of v_L is due to the tendency of an inductance to oppose a change in current. Assuming that $\alpha L > R$, the total voltage is negative as shown in Fig. 5.1b.

The impedance $Z(s) = R + sL$ and, therefore, the response $v(t)$ vary widely with changes in parameter s. For a given magnitude I_0, the inductive voltage v_L is directly proportional to s and changes sign with s. For the special case of $s = -R/L$, $Z(s) = 0$, or the circuit presents no opposition to this particular current function. Our interpretation is that, since current is possible with no forcing voltage, this is a natural response current. Two other interesting cases are for $s = 0$ and $s = j\omega$.

Response to Direct Currents

A unidirectional current of constant magnitude is called a continuous or direct current, abbreviated dc. In common usage a voltage of the same form is called a *dc voltage*, and we sometimes talk of *dc currents* in contrast to alternating or *ac currents*. Direct-current energy supplied by chemical batteries, thermocouples, solar cells, or rotating dc generators is used for a wide variety of purposes including ship propulsion, electrolytic refining, electron beam acceleration, and transistor operation. In circuit diagrams dc voltage sources appear so often that the special symbol of Fig. 5.2 is used; the longer line indicates the positive terminal.

For the special case of $s = 0$ (see Fig. 4.12), the general exponential becomes

$$i = I_0\, e^{0t} = I_{\mathrm{dc}} = I\dagger \tag{5-2}$$

The dc values of impedance $Z(s)$, for the case of $s = 0$, we designate $Z(0)$. For resistance, inductance, and capacitance these impedances are:

$$Z_R(0) = R \quad Z_L(0) = 0 \quad Z_C(0) = \infty \tag{5-3}$$

The same results are obtained by considering the fundamental element equations. In a resistance, $v = Ri$ for all waveforms; for a direct current I, a direct voltage $V = RI$ appears across a resistance. The impedance $v/i = V/I = R$ as expected.

When a direct current $i = I$ flows in an inductance, $di/dt = 0$ and, therefore, $v = L(di/dt) = 0$. The impedance $v/i = 0$, or an inductance has the characteristic of a short circuit. We say "an inductance looks like a short circuit to

---

† Capital I and V are used to designate constant values, such as phasors or dc magnitudes, and lower case i and v to designate functions of time.

a dc current." When a direct voltage is applied to an inductance carrying no initial current,

$$i = \frac{1}{L}\int_0^t v\,dt = \frac{V}{L}t \tag{5-4}$$

and the current increases linearly with time.

When a direct voltage $v = V$ appears across a capacitance, $dv/dt = 0$ and, therefore, $i = C(dv/dt) = 0$. The impedance $v/i = \infty$ or we say, "a capacitance looks like an open circuit to a dc voltage." When a direct current $i = I$ flows in an initially uncharged capacitance,

$$v = \frac{1}{C}\int_0^t i\,dt = \frac{I}{C}t \tag{5-5}$$

and the voltage increases linearly with time.

Example 1

Switch S in Fig. 5.2a has been closed for a long time. Determine the current i and the voltages across R_1, R_2, and R_3.

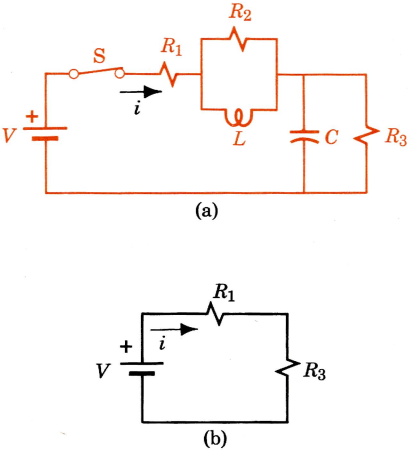


Figure 5.2 Forced response to a dc voltage.

We assume that after the "long time" specified all natural response has died away and only a forced response exists. For a direct forcing voltage the response is a direct current. To a direct current, inductance L looks like a short circuit effectively removing R_2 from the circuit; therefore, the voltage across R_2 is zero.

To a direct voltage, capacitance C looks like an open circuit and it can be removed from the circuit; therefore, all the current i flows through R_3. For the dc case, the circuit looks like Fig. 5.2b.

The total series resistance is $R_1 + R_3$ and

$$i = I = \frac{V}{R_1 + R_3}$$

Hence,

$$v_1 = R_1 i = \frac{R_1}{R_1 + R_3}V$$

and

$$v_3 = R_3 i = \frac{R_3}{R_1 + R_3}V$$

Response to Sinusoids

The representation of sinusoids by exponentials leads to the phasor concept whereby functions of time are transformed into constant quantities for easy manipulation and display. Also, we know that in dealing with exponentials the impedance concept provides a convenient method for finding forced response. Therefore, we expect exponential representation of sinusoids to lead to a convenient method for analyzing circuits with sinusoidal sources and for displaying the results.

To confirm our expectation, let us consider the series *RLC* circuit of Fig. 5.3 where $i = I_m \cos \omega t$. Applying Kirchhoff's voltage law to obtain the connection equation,

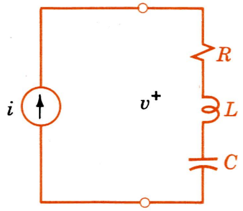


Figure 5.3 A series *RLC* circuit.

$$v = Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt \tag{5-6}$$

Under steady-state conditions, v is the forced response and will be sinusoidal for a sinusoidal forcing current. Anticipating the result, let us assume that voltage $v = \text{Re}\,\{A\,e^{j\omega t}\}$ where constant $A = V_m\,e^{j\theta} = \mathbf{V}_m$ (a phasor). Letting

$$i = I_m \cos(\omega t + 0) = \text{Re}\,\{I_m\,e^{j0}\,e^{j\omega t}\} = \text{Re}\,\{\mathbf{I}_m\,e^{j\omega t}\}$$

the terms in Eq. 5-6 are:

$$Ri = R\,\text{Re}\,\{\mathbf{I}_m\,e^{j\omega t}\} = \text{Re}\,\{R\mathbf{I}_m\,e^{j\omega t}\}$$

$$L\frac{di}{dt} = L\frac{d}{dt}\text{Re}\,\{\mathbf{I}_m\,e^{j\omega t}\} = L\,\text{Re}\,\{\mathbf{I}_m\,(j\omega)\,e^{j\omega t}\} = \text{Re}\,\{j\omega L\mathbf{I}_m\,e^{j\omega t}\}$$

and†

$$\frac{1}{C}\int i\,dt = \frac{1}{C}\int \text{Re}\,\{\mathbf{I}_m\,e^{j\omega t}\}\,dt = \frac{1}{C}\text{Re}\left\{\frac{\mathbf{I}_m}{j\omega}e^{j\omega t}\right\} = \text{Re}\left\{\frac{1}{j\omega C}\mathbf{I}_m\,e^{j\omega t}\right\}$$

Substituting in Eq. 5-6,

$$\text{Re}\,\{\mathbf{V}_m\,e^{j\omega t}\} = \text{Re}\,\{R\mathbf{I}_m\,e^{j\omega t}\} + \text{Re}\,\{j\omega L\mathbf{I}_m\,e^{j\omega t}\} + \text{Re}\left\{\frac{1}{j\omega C}\mathbf{I}_m\,e^{j\omega t}\right\}$$

Dividing through by $\sqrt{2}$ to convert all phasors to effective values, and following the rule for the addition of complex quantities,

$$\text{Re}\,\{\mathbf{V}\,e^{j\omega t}\} = \text{Re}\left\{\left(R\mathbf{I} + j\omega L\mathbf{I} + \frac{1}{j\omega C}\mathbf{I}\right)e^{j\omega t}\right\}$$

By the rule for equality, if the real parts of these two complex quantities are to be equal at all times (assuming that t is the only variable), then the complex quantities themselves are equal. This requires that

$$\mathbf{V} = R\mathbf{I} + j\omega L\mathbf{I} + \frac{1}{j\omega C}\mathbf{I} = \left(R + j\omega L + \frac{1}{j\omega C}\right)\mathbf{I} \tag{5-7}$$

---

† It can be shown that, with respect to a real variable such as time,

$$\frac{d}{dt}\text{Re}\,\{\mathbf{W}\} = \text{Re}\left\{\frac{d\mathbf{W}}{dt}\right\} \quad \text{and} \quad \int \text{Re}\,\{\mathbf{W}\}\,dt = \text{Re}\left\{\int \mathbf{W}\,dt\right\}$$

This result is reassuring because the quantity in parentheses looks just like $Z(s) = R + sL + 1/sC$ for a series *RLC* circuit if s has the value $j\omega$.

From Eq. 5-7 we conclude that phasor voltage and phasor current are related by a *complex impedance* $\mathbf{Z} = Z(j\omega)$ or

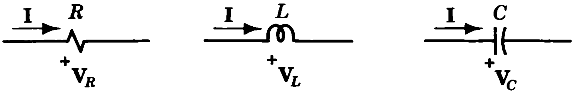
$$\mathbf{V} = \mathbf{ZI} \tag{5-8}$$

which is sometimes referred to as the Ohm's law of sinusoidal circuits.

Impedance $\mathbf{Z}$ is measured in ohms when phasors $\mathbf{V}$ and $\mathbf{I}$ are in volts and amperes, respectively. Note that $\mathbf{Z}$ is a complex quantity, like a phasor, but $\mathbf{Z}$ is *not* a phasor. The term phasor is reserved for quantities representing sinusoidal varying functions of time. The chief virtue of complex impedance is that although it expresses the relation between two time-varying quantities, $\mathbf{Z}$ itself is *not* a function of time.

Phasor Methods

The element equations define the voltage-current characteristics of passive elements. From Eq. 5-7 we can form the set of steady-state relations shown in Fig. 5.4. The sign convention indicates that for the assumed reference direction of positive current the corresponding reference polarity for voltage is as shown. Such a convention is necessary because the actual current and voltage are both changing sign periodically, but not simultaneously.



$$\mathbf{V}_R = R\mathbf{I} \qquad \mathbf{V}_L = j\omega L\mathbf{I} \qquad \mathbf{V}_C = \frac{1}{j\omega C}\mathbf{I}$$

$$\mathbf{I}_R = \frac{1}{R}\mathbf{V} \qquad \mathbf{I}_L = \frac{1}{j\omega L}\mathbf{V} \qquad \mathbf{I}_C = j\omega C\,\mathbf{V}_C$$

Figure 5.4 Element equations in terms of phasors.

Circuit Laws in Terms of Phasors

The connection equations for sinusoidal forced response are based on Kirchhoff's laws. Assuming that all sources in a linear circuit are sinusoids of the same frequency, all steady-state voltages and currents are sinusoids of that same frequency. Kirchhoff's voltage law

$$\Sigma v = 0 = v_1 + v_2 + \cdots + v_k$$

becomes, in terms of phasors,

$$\Sigma v = 0 = \mathrm{Re}\,\{\mathbf{V}_1\, e^{j\omega t}\} + \mathrm{Re}\,\{\mathbf{V}_2\, e^{j\omega t}\} + \cdots + \mathrm{Re}\,\{\mathbf{V}_k\, e^{j\omega t}\} \qquad (5\text{-}9)$$

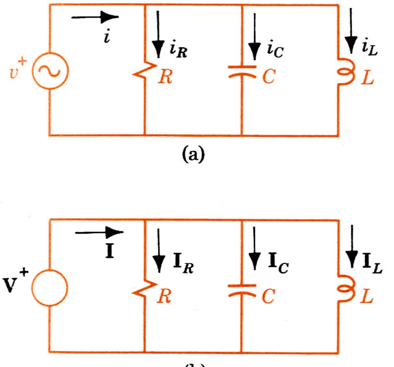
By the rule of equality, if Eq. 5-9 is to hold for all values of t,

$$0 = \mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_k = \Sigma\mathbf{V} \qquad (5\text{-}10)$$

In other words, for steady-state sinusoidal functions, Kirchhoff's voltage law holds for voltage phasors and $\Sigma\mathbf{V} = 0$ around any closed loop. Following a similar line of reasoning, Kirchhoff's current law holds for current phasors and $\Sigma\mathbf{I} = 0$ into any node.

Example 2

Voltage $v = 120\sqrt{2}\cos(1000t + 90)$ V is applied to the circuit of Fig. 5.5a where $R = 15\,\Omega$, $C = 83.3\,\mu\text{F}$, and $L = 30$ mH. Find $i(t)$.



(a)

(b)

Figure 5.5 Forced response to a sinusoidal voltage.

The first step is to transform $v(t)$ into phasor $\mathbf{V} = 120\underline{/90^\circ}$ V. Then

$$\mathbf{I}_R = \frac{1}{R}\mathbf{V} = \frac{1}{15}120\underline{/90^\circ} = 8\underline{/90^\circ} = 0 + j8\text{ A}$$

$$\begin{aligned}\mathbf{I}_C = j\omega C\mathbf{V} &= (0.0833\underline{/90^\circ})(120\underline{/90^\circ})\\ &= 10\underline{/180^\circ} = -10 + j0\text{ A}\end{aligned}$$

$$\mathbf{I}_L = \frac{1}{j\omega L}\mathbf{V} = \frac{120\underline{/90^\circ}}{30\underline{/90^\circ}} = 4\underline{/0^\circ} = 4 + j0\text{ A}$$

By Kirchhoff's current law, $\Sigma\mathbf{I} = 0$ or

$$\begin{aligned}\mathbf{I} = \mathbf{I}_R + \mathbf{I}_C + \mathbf{I}_L &= (0 - 10 + 4) + j(8 + 0 + 0)\\ &= -6 + j8 = 10\underline{/127^\circ}\text{ A}\end{aligned}$$

The final step is to transform phasor $\mathbf{I}$ into

$$i(t) = 10\sqrt{2}\cos(1000t + 127^\circ)\text{ A}$$

AC Impedance and Admittance

Although an "alternating" current can be nonsinusoidal, unless otherwise stated we assume that an alternating current or an "ac voltage" is sinusoidal. Most electrical energy is generated, transmitted, and consumed in the form of alternating current. Also, much of the dc energy used in power and communication applications is first generated as ac energy and then converted to dc form. The number of applications of ac circuits is so great that we can justify taking time to learn some short cuts in solving such circuits.

Taking another look at Eq. 5-7, repeated here for convenience,

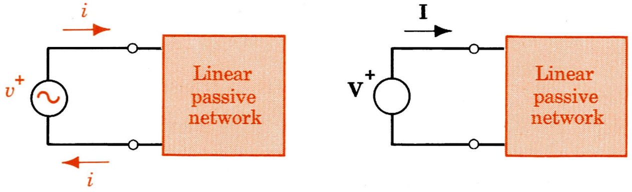
$$\mathbf{V} = R\mathbf{I} + j\omega L\mathbf{I} + \frac{1}{j\omega C}\mathbf{I} = \left(R + j\omega L + \frac{1}{j\omega C}\right)\mathbf{I} \tag{5-7}$$

we see that the total complex impedance of a series RLC circuit consists of the sum of the impedances of the elements. Just as each passive circuit element has associated with it an impedance to exponentials $Z(s)$ and an impedance to direct currents $Z(0)$, it also has an impedance to sinusoids $Z(j\omega)$.

$$\begin{aligned} &\text{For resistance,} && Z_R(j\omega) = \mathbf{Z}_R = R = R\underline{/0^\circ} \\ &\text{For inductance,} && Z_L(j\omega) = \mathbf{Z}_L = j\omega L = \omega L\underline{/90^\circ} \\ &\text{For capacitance,} && Z_C(j\omega) = \mathbf{Z}_C = \frac{1}{j\omega C} = j\left(-\frac{1}{\omega C}\right) = \left(\frac{1}{\omega C}\right)\underline{/-90^\circ} \end{aligned} \tag{5-11}$$

The quantity R is called the *ac resistance*† and is measured in ohms. The corresponding term for an inductance is called the *inductive reactance* $X_L = \omega L$ in ohms. For a capacitance, the corresponding term is called *capacitive reactance* $X_C = -1/\omega C$ in ohms. The significance of the minus sign is indicated in Fig. 5.10.

In Fig. 5.6a, a linear passive network is energized by a voltage source applied to two terminals brought out for the purpose. The current out of the lower terminal is just equal to the current into the upper terminal. (Why?) Such a *two-terminal* network can be represented by Fig. 5.6b if voltage v is sinusoidal.



(a) Time functions (b) Phasors

Figure 5.6 Representation of a two-terminal network.

The steady-state sinusoidal response of any two-terminal network is completely defined by the impedance $\mathbf{Z} = \mathbf{V}/\mathbf{I}$ at the terminals. Another way of characterizing a network is in terms of the admittance $Y(s)$, defined in Chapter 4 as the ratio of exponential current to voltage. For sinusoidal current and voltage,

$$Y(j\omega) = \mathbf{Y} = \frac{\mathbf{I}}{\mathbf{V}}$$

where $\mathbf{Y}$ is the complex admittance in mhos.

In general $\mathbf{Z}$ and $\mathbf{Y}$ are complex and can be written as

---

† The ac resistance of a physical resistor may be different from the dc resistance.

$$\mathbf{Z} = Z\underline{/\phi_Z} = R + jX \quad \text{and} \quad \mathbf{Y} = Y\underline{/\phi_Y} = G + jB \tag{5-12}$$

The real part of **Z** is called *resistance* in ohms and is designated R; in the RLC circuit described by Eq. 5-7, the real part of **Z** is the resistance of a particular element, but in general this is *not* true. The imaginary part of **Z** is called *reactance* in ohms and is designated X; for the series RLC circuit,

$$X = X_L + X_C = \omega L - 1/\omega C$$

The real part of **Y** is called *conductance* in mhos and is designated G; in general, G is *not* the conductance of a particular element. The imaginary part of **Y** is called *susceptance* in mhos and is designated B. A positive B is associated with a predominantly capacitive circuit.

Example 3

The circuit of Example 2 is repeated in Fig. 5.7, where $\mathbf{V} = 120\underline{/90^\circ}$ V, $\omega = 1000$ rad/s, $R = 15\ \Omega$, $C = 83.3\ \mu$F, and $L = 30$ mH; find **I**.

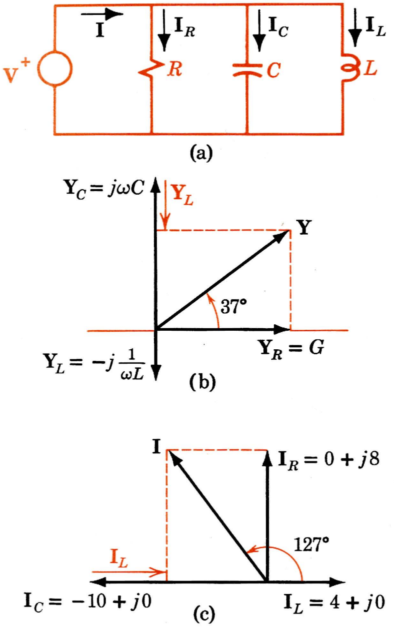


Figure 5.7 Use of the complex admittance concept.

The admittance concept is particularly useful in dealing with parallel circuits since admittances of parallel elements are added directly. From the element equations displayed in Fig. 5.4,

$$\mathbf{Y}_R = \frac{\mathbf{I}}{\mathbf{V}} = \frac{1}{R} = \frac{1}{15} = 0.0667\ \mho$$

$$\mathbf{Y}_C = j\omega C = j1000 \times 83.3 \times 10^{-6} = j0.0833\ \mho$$

$$\mathbf{Y}_L = \frac{1}{j\omega L} = j\frac{-1}{1000 \times 0.03} = -j0.0333\ \mho$$

The admittance at the terminals of the one-port is

$$\begin{aligned}\mathbf{Y} = \mathbf{Y}_R + \mathbf{Y}_C + \mathbf{Y}_L &= 0.0667 + j(0.0833 - 0.0333)\\ &= 0.0667 + j0.05\ \mho\end{aligned}$$

In polar form,

$$\begin{aligned}\mathbf{Y} &= \sqrt{(0.0667)^2 + (0.05)^2}\ \underline{/\arctan\frac{0.05}{0.0667}}\\ &= 0.0833\underline{/37^\circ}\ \mho\end{aligned}$$

and

$$\begin{aligned}\mathbf{I} = \mathbf{YV} &= (0.0833\underline{/37^\circ})(120\underline{/90^\circ})\\ &= 10\underline{/127^\circ}\ \text{A}\end{aligned}$$

The components of **Y** are shown in Fig. 5.7b; for comparison the component currents calculated in Example 2 are shown in Fig. 5.7c.

Since **Z** and **Y** are complex quantities, they obey the rules of complex algebra and can be represented by plane vectors. In contrast to phasors, **Z** and **Y** can lie only in the first and fourth quadrants of the complex plane because R and G are always positive for passive networks. (See Example 3.) Since $\mathbf{Z} = 1/\mathbf{Y}$,

$$R + jX = \frac{1}{G + jB} = \frac{G}{G^2 + B^2} - j\frac{B}{G^2 + B^2} \tag{5-13}$$

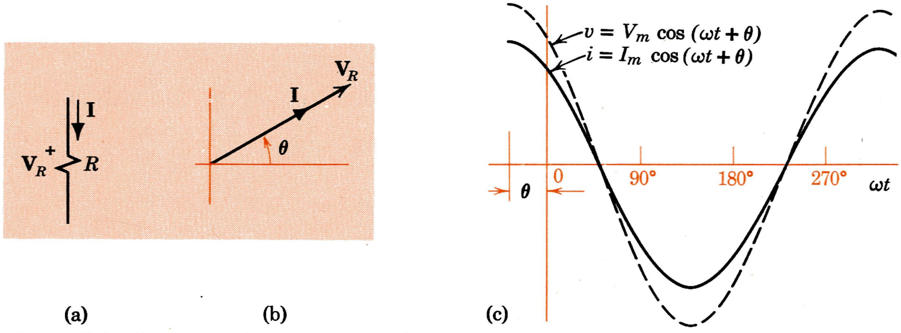
Similarly,

$$G + jB = \frac{1}{R + jX} = \frac{R}{R^2 + X^2} - j\frac{X}{R^2 + X^2} \tag{5-14}$$

It must be emphasized that R and G, and X and B are the real and imaginary parts of quantities defined at the terminals and, in general, are not associated with particular elements of a network. (This point is illustrated in Example 4, p. 129.) Failure to realize this is sure to result in mistakes in the solution of complicated circuits.

Voltage and Current Relations

By definition, impedance $Z(j\omega)$ is the ratio of the voltage phasor to the current phasor or $\mathbf{Z} = \mathbf{V}/\mathbf{I}$ where all quantities are complex. For a resistance (Fig. 5.8), $\mathbf{Z}_R = \mathbf{V}_R/\mathbf{I} = R\underline{/0^\circ}$; therefore, the voltage and current phasors are drawn at the same angle. The voltage and current are shown as functions of time in Fig. 5.8c; the voltage and current waves reach maximum and zero values simultaneously. We say:

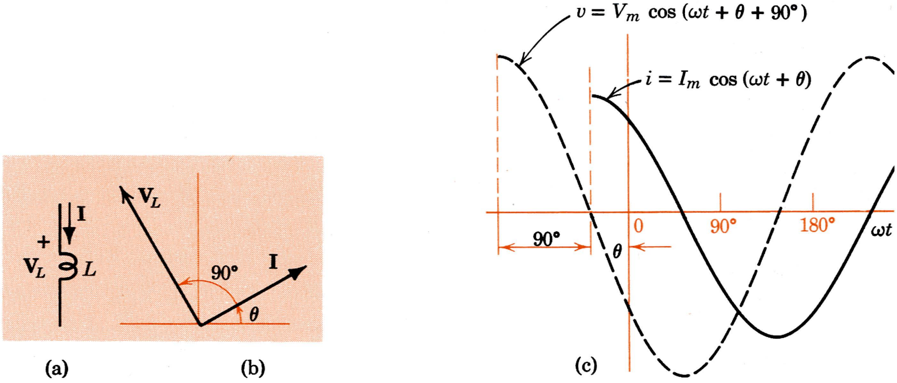

I
+
$\mathbf{V}_R$
R
$\mathbf{V}_R$
I
θ
$v = V_m \cos(\omega t + \theta)$
$i = I_m \cos(\omega t + \theta)$
0
90°
180°
270°
ωt
θ

(a) (b) (c)

Figure 5.8 Current-voltage relations for a resistance.

The voltage across a resistance is in phase with the current through it.

For an inductance (Fig. 5.9), $\mathbf{Z}_L = \mathbf{V}_L/\mathbf{I} = \omega L\underline{/90^\circ}$; therefore, the voltage phasor is 90° *ahead* of the current phasor. When the voltage and current are shown as functions of time (Fig. 5.9c), it is clear that the voltage wave reaches


$v = V_m \cos(\omega t + \theta + 90°)$
$i = I_m \cos(\omega t + \theta)$
I
+
V_L
L
V_L
90°
I
θ
90°
θ
0
90°
180°
ωt

(a) (b) (c)

Figure 5.9 Current-voltage relations for an inductance.

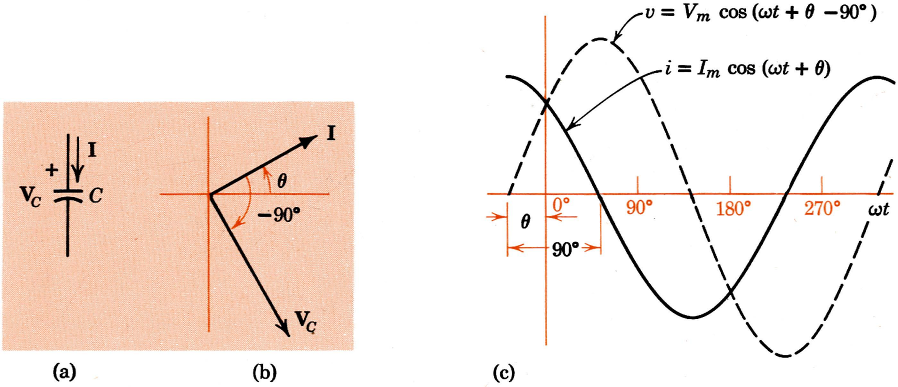
its maximum and zero values at an *earlier* time than the current wave. We say:

> *The voltage across an inductive reactance leads by* 90° *the current through it.*

For a capacitance (Fig. 5.10), $\mathbf{Z}_C = \mathbf{V}_C/\mathbf{I}_C = (1/\omega C)\underline{/-90°}$; therefore the voltage phasor is 90° *behind* the current phasor. When the voltage and current are shown as functions of time (Fig. 5.10c), it is clear that the voltage wave reaches its maximum and zero values at a *later* time than the current wave. We say:

> *The voltage across a capacitive reactance lags by* 90° *the current through it.*

These phrases emphasize that the *voltage across an element* is specified with respect to the *current through that element* and not with respect to any other current.


$v = V_m \cos(\omega t + \theta - 90°)$
$i = I_m \cos(\omega t + \theta)$
I
+
V_C
C
I
θ
−90°
V_C
θ
0°
90°
90°
180°
270°
ωt

(a) (b) (c)

Figure 5.10 Current-voltage relations for a capacitance.

AC Circuit Analysis

By using the impedance concept, differential equations are replaced by algebraic equations. By using the phasor concept, continuously varying quantities are replaced by complex constants. With these aids the analysis of ac circuits becomes a relatively simple process.

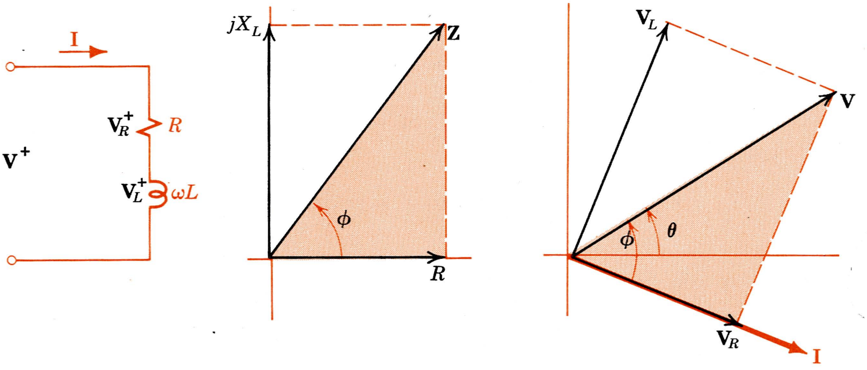
Series RL and RC Circuits

The process is illustrated by the analysis of the circuit in Fig. 5.11. For the series combination of R and L,

$$\mathbf{Z} = R + jX_L = R + j\omega L = Z\underline{/\phi} = \sqrt{R^2 + (\omega L)^2}\ \underline{/\arctan \frac{\omega L}{R}} \tag{5-15}$$

Assuming that $\mathbf{V}$ is given as $V\underline{/\theta} = (V_m/\sqrt{2})\underline{/\theta}$,

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{V\underline{/\theta}}{Z\underline{/\phi}} = \frac{V}{Z}\underline{/\theta - \phi} = I\underline{/\theta - \phi} \tag{5-16}$$



(a) Diagram (b) Complex impedance (c) Phasor diagram

Figure 5.11 Series *RL* circuit analysis.

As shown in Fig. 5.11c, the total voltage phasor leads the current phasor† by *phase angle* ϕ, something less than 90°. The voltage across the resistance is in phase with the current, and the voltage across the inductive reactance is 90° ahead of the current. These are components of the total applied voltage, and the phasor sum of $\mathbf{V}_R$ and $\mathbf{V}_L$ is equal to $\mathbf{V}$.

---

† In drawing phasor diagrams, all voltages are drawn to one scale and all currents are drawn to one scale (which may be quite different from the voltage scale).

For the series combination of R and C (Fig. 5.12a),

$$\mathbf{Z} = R + jX_C$$

$$= R + j\left(\frac{-1}{\omega C}\right) = Z\underline{/\phi} = \sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}\ \underline{\Big/\arctan\frac{-1/\omega C}{R}} \qquad (5\text{-}17)$$

where ϕ is a negative angle. Assuming that $\mathbf{I}$ is given as $I\underline{/\theta} = (I_m/\sqrt{2})\underline{/\theta}$,

$$\mathbf{V} = \mathbf{ZI} = (Z\underline{/\phi})(I\underline{/\theta}) = ZI\underline{/\phi + \theta} = V\underline{/\phi + \theta} \qquad (5\text{-}18)$$

As shown in Fig. 5.12c, the voltage across the resistance R is in phase with the current $\mathbf{I}$. The voltage across the capacitive reactance X_C lags behind the current by 90°. The resulting total voltage lags behind the current by an angle less than 90°.

Example 4

In the circuit of Fig. 5.12a below, $\mathbf{V} = 20\underline{/45°}$ V, $R = 5\ \Omega$, and reactance $X_C = -1/\omega C = -8.66\ \Omega$. Find $\mathbf{I}$, $\mathbf{V}_R$, and $\mathbf{V}_C$.

First, we calculate $\mathbf{Z}$ and $\mathbf{Y}$.
By Eq. 5-17,

$$\mathbf{Z} = R + j\left(\frac{-1}{\omega C}\right) = 5 - j8.66 = 10\underline{/-60°}\ \Omega$$

$$\mathbf{Y} = \frac{1}{\mathbf{Z}} = 0.1\underline{/+60°} = 0.05 + j0.0866\ \Omega$$

Note that $G = 0.05\ ℧ = R/(R^2 + X^2)$ and *not* $1/R$.
Then we calculate $\mathbf{I}$.
By Eq. 5-8,

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{20\underline{/45°}}{10\underline{/-60°}} = 2\underline{/105°}\ \text{A}$$

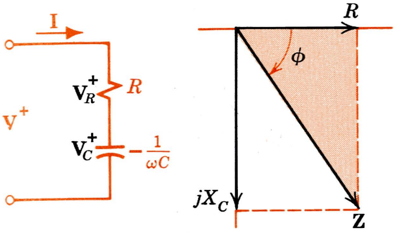
Hence,

$$\mathbf{V}_R = R\mathbf{I} = (5\underline{/0°})(2\underline{/105°}) = 10\underline{/105°}\ \text{V}$$

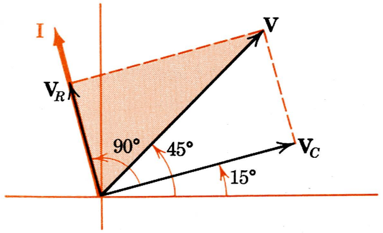
and

$$\mathbf{V}_C = \mathbf{Z}_C\mathbf{I} = (8.66\underline{/-90°})(2\underline{/105°})$$
$$= 17.32\underline{/15°}\ \text{V}$$

These results are plotted in Fig. 5.12c.



(a) Diagram (b) Complex impedance



(c) Phasor diagram

Figure 5.12 Series RC circuit analysis.

Assuming ω is known, the phasors in Example 4 could be transformed into time functions, but this is seldom necessary. For the reasons discussed in Chapter 3, the quantities of interest in steady-state circuit analysis are the effective phasors and ammeters and voltmeters are calibrated to read these values directly.

Parallel GL and GC Circuits

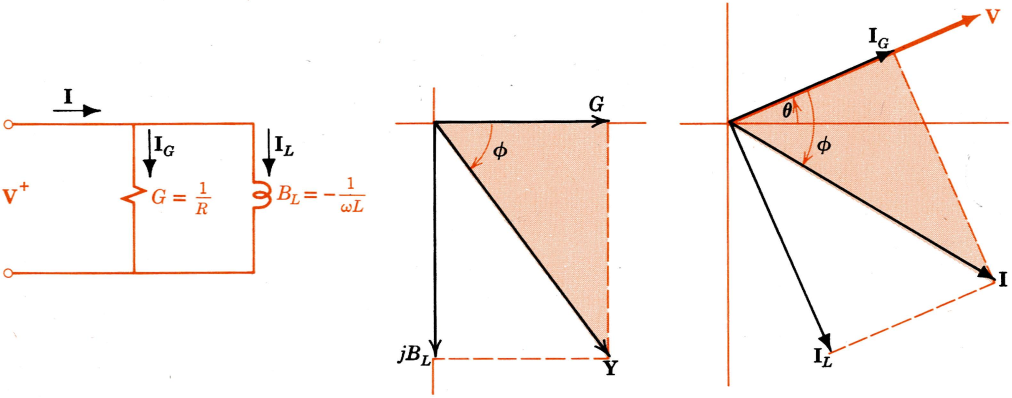
It was pointed out earlier that in dealing with parallel circuits the admittance approach is particularly useful. The process is illustrated by the analysis of the circuit in Fig. 5.13a. For the parallel combination of G and L,

$$\mathbf{Y} = G + jB_L = G + j\left(\frac{-1}{\omega L}\right) = Y\underline{/\phi} = \sqrt{G^2 + \left(\frac{1}{\omega L}\right)^2}\;\underline{/\arctan\frac{-1/\omega L}{G}} \tag{5-19}$$

where ϕ is a negative angle. Assuming that $\mathbf{V}$ is given as $V\underline{/\theta}$,

$$\mathbf{I} = \mathbf{YV} = (Y\underline{/\phi})(V\underline{/\theta}) = YV\underline{/\phi + \theta} = I\underline{/\phi + \theta} \tag{5-20}$$

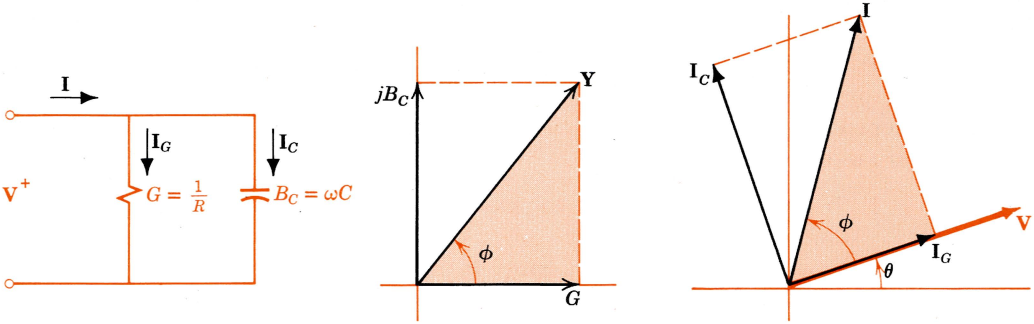
As shown in Fig. 5.13c, the current in the conductance G is in phase with the voltage $\mathbf{V}$. The current in the inductive susceptance B_L lags behind the voltage by 90°. The resulting total current I lags behind the voltage by an angle ϕ less than 90°.



(a) Diagram (b) Complex admittance (c) Phasor diagram

Figure 5.13 Parallel GL circuit analysis.

For the parallel combination of G and C (Fig. 5.14),



(a) Diagram (b) Complex admittance (c) Phasor diagram

Figure 5.14 Parallel *GC* circuit analysis.

$$\mathbf{Y} = G + jB_C = G + j\omega C = Y\angle\phi = \sqrt{G^2 + (\omega C)^2}\ \Big/ \arctan\frac{\omega C}{G} \quad \text{(5-21)}$$

Equation 5-20 applies for this circuit as well, but here phase angle ϕ is positive and the total current phasor *leads* the voltage by something less than 90°.

Although the admittance approach is usually easier, parallel circuits can also be analyzed in terms of impedances properly combined. In general, for parallel elements,

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n \quad \text{(5-22)}$$

or

$$\frac{1}{\mathbf{Z}} = \frac{1}{\mathbf{Z}_1} + \frac{1}{\mathbf{Z}_2} + \cdots + \frac{1}{\mathbf{Z}_n} \quad \text{(5-23)}$$

For the special case of two parallel elements,

$$\frac{1}{\mathbf{Z}} = \frac{1}{\mathbf{Z}_1} + \frac{1}{\mathbf{Z}_2} \quad \text{or} \quad \mathbf{Z} = \frac{\mathbf{Z}_1\mathbf{Z}_2}{\mathbf{Z}_1 + \mathbf{Z}_2} \quad \text{(5-24)}$$

For a parallel *GL* circuit, the impedance at the terminals is

$$\mathbf{Z} = \frac{\mathbf{Z}_G\mathbf{Z}_L}{\mathbf{Z}_G + \mathbf{Z}_L} = \frac{R(j\omega L)}{R + j\omega L}$$

Then

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{R + j\omega L}{jR\omega L}\mathbf{V} = \left(\frac{1}{j\omega L} + \frac{1}{R}\right)\mathbf{V} = \left(G + j\frac{-1}{\omega L}\right)\mathbf{V}$$

which agrees with Eqs. 5-19 and 5-20.

In talking about circuit analysis in general, we need a term that includes both impedance and admittance. For this purpose we shall use the word *immittance* derived from *im*pedance and ad*mittance*.

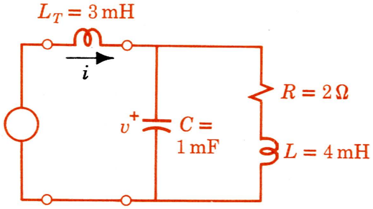
General Circuit Analysis

We are now ready to formulate a general procedure applicable to a variety of problems in steady-state sinusoidal circuits. Although the exact sequence depends on the particular problem (see Examples 5 and 6), the following steps are usually necessary:

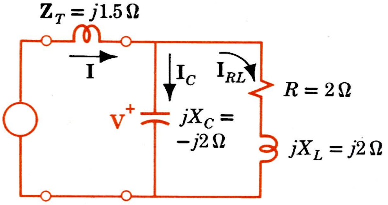
1. Transform time functions to phasors and convert element values to complex immittances.
2. Combine immittances in series or parallel to simplify the circuit.
3. Determine the desired response in phasor form, using connection equations.
4. Draw a phasor diagram to check computations and to display results.
5. Transform phasors to time functions if required.

Example 5

The circuit of Fig. 5.15a represents a "load" consisting of C, R, and L supplied by a generator over a transmission line approximated by L_T. For a load voltage $v = 400 \cos (500t + 45°)$ V, determine the current $i(t)$.



(a) Element values



(b) Complex impedances

Figure 5.15 General circuit analysis.

In phasor form, $v(t)$ becomes $\mathbf{V} = 283\underline{/45^\circ}$ V.

$$\mathbf{Z}_T = j\omega L_T = j500 \times 3 \times 10^{-3} = j1.5\,\Omega$$

$$\mathbf{Y}_C = j\omega C = j500 \times 10^{-3} = j0.5\,℧$$

$$\mathbf{Z}_C = 1/\mathbf{Y}_C = -j2\,\Omega$$

$$\mathbf{Z}_L = j\omega L = j500 \times 4 \times 10^{-3} = j2\,\Omega$$

The circuit diagram of Fig. 5.15b displays these values. Since $\mathbf{I} = \mathbf{YV}$, we need the admittance of the parallel circuit constituting the load.

$$\begin{aligned} \mathbf{Y}_{RL} &= \frac{1}{\mathbf{Z}_{RL}} = \frac{1}{2 + j2} = \frac{1}{2\sqrt{2}\underline{/45^\circ}} \\ &= 0.25\sqrt{2}\underline{/-45^\circ} = 0.25 - j0.25\,℧ \end{aligned}$$

$$\begin{aligned} \mathbf{Y} &= \mathbf{Y}_C + \mathbf{Y}_{RL} = j0.5 + (0.25 - j0.25) \\ &= 0.25 + j0.25 = 0.25\sqrt{2}\underline{/+45^\circ}\,℧ \end{aligned}$$

Then

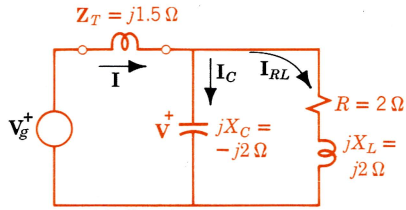
$$\mathbf{I} = \mathbf{YV} = (0.25\sqrt{2}\underline{/45^\circ})(283\underline{/45^\circ}) = 100\underline{/90^\circ}\text{ A}$$

as shown in Fig. 5.16b of Example 6, p. 133 and

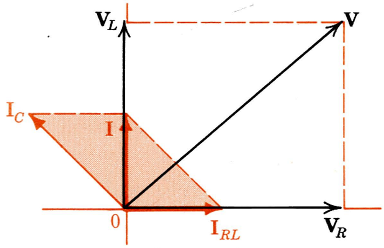
$$i(t) = 100\sqrt{2} \cos (500t + 90°)\text{ A}$$

Example 6

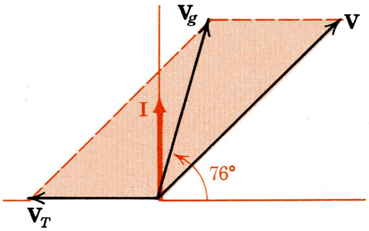
For the circuit of Example 5 (repeated in Fig. 5.16a), determine the generator voltage required at the other end of the transmission line to provide the specified load voltage.



(a) Circuit impedances



(b) Current phasors



(c) Voltage phasors

Figure 5.16 Use of phasor diagrams.

The phasor diagram should be drawn as the calculations proceed, to provide a continuous check as well as a final display of results. In addition to the given V and the calculated I we know that:

$$\mathbf{I}_C = \mathbf{Y}_C\mathbf{V} = (0.5\angle 90^\circ)(283\angle 45^\circ) = 141\angle 135^\circ \text{ A}$$

$$\mathbf{I}_{RL} = \mathbf{Y}_{RL}\mathbf{V} = (0.35\angle -45^\circ)(283\angle 45^\circ) = 100\angle 0^\circ \text{ A}$$

$$\mathbf{V}_R = R\mathbf{I}_{RL} = (2\angle 0^\circ)(100\angle 0^\circ) = 200\angle 0^\circ \text{ V}$$

$$\mathbf{V}_L = \mathbf{Z}_L\mathbf{I}_{RL} = (2\angle 90^\circ)(100\angle 0^\circ) = 200\angle 90^\circ \text{ V}$$

These are shown on the phasor diagram of Fig. 5.16b.

The construction lines indicate consistency with the connection equations:

$$\mathbf{I} = \mathbf{I}_C + \mathbf{I}_{RL} \quad \text{and} \quad \mathbf{V} = \mathbf{V}_R + \mathbf{V}_L.$$

Other checks are also available: $\mathbf{I}_C$ leads $\mathbf{V}$ by 90°, $\mathbf{V}_R$ is in phase with $\mathbf{I}_{RL}$, and $\mathbf{V}_L$ leads $\mathbf{I}_{RL}$ by 90°.

To determine the generator voltage, we write a loop equation, $\Sigma\mathbf{V} = 0 = \mathbf{V}_g - \mathbf{I}\mathbf{Z}_T - \mathbf{V}$ and solve for

$$\begin{aligned}\mathbf{V}_g &= \mathbf{Z}_T\mathbf{I} + \mathbf{V} = (1.5\angle 90^\circ)(100\angle 90^\circ) + 283\angle 45^\circ \\ &= 150\angle 180^\circ + 283\angle 45^\circ \\ &= -150 + j0 + 200 + j200 \\ &= 50 + j200 = 206/76^\circ \text{ V}\end{aligned}$$

This result is confirmed by the phasor diagram of Fig. 5.16c.

The final step, if required, is to transform phasors into functions of time. By inspection,

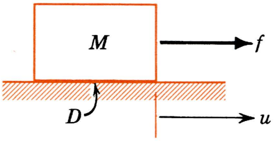
$$v_g = 206\sqrt{2}\cos(500t + 76^\circ) \text{ V}$$

Analogs and Duals

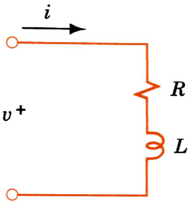
One justification for this intensive study of electrical circuit analysis is that the techniques developed are equally applicable to analogous situations in other fields. By definition:

> *Analogous systems are described by the same integro-differential equation or set of equations.*

The solution of the equation describing one physical system automatically yields the solution for any analogous system. The solution of an electrical circuit provides results that are directly applicable to analogous mechanical, hydraulic, or thermal circuits.



(a) Mechanical



(b) Electrical

Figure 5.17 Examples of analogous circuits.

Analogs

As an illustration, consider the mass M sliding at velocity u with a friction coefficient D on a surface as shown in Fig. 5.17a (same as Fig. 4.3). By d'Alembert's principle, the governing equation is

$$Du + M\frac{du}{dt} = f \tag{5-25}$$

This looks familiar; in form it is similar to the equation resulting from the application of Kirchhoff's voltage law to the circuit of Fig. 4.1 (repeated here as Fig. 5.17b)

$$Ri + L\frac{di}{dt} = v \tag{5-26}$$

Mathematically, these two integro-differential equations are the same and, therefore, the circuits are analogous; the solutions of the two equations are identical if simple symbol substitutions are permitted.

For exponential functions we expect a *motional impedance* $Z(s) = D + sM$. The natural behavior is defined by $Z(s) = 0$ or $s = -D/M$ and $u = U_0\, e^{-(D/M)t}$. A steady pull $f = F$ on the mass corresponds to a "direct" force; then $Z(0) = D$, and $U = F/D$ is the forced response to a direct force.

For a sinusoidal pull, $f = \sqrt{2}\, F \cos(\omega t + \theta)$, a phasor approach is indicated where $\mathbf{F} = F\angle\theta$. The motional impedance is

$$Z(j\omega) = \mathbf{Z} = D + j\omega M = Z\angle\phi \tag{5-27}$$

where

$$Z = \sqrt{D^2 + (\omega M)^2} \quad \text{and} \quad \phi = \arctan\frac{\omega M}{D} \tag{5-28}$$

Then

$$\mathbf{U} = \frac{\mathbf{F}}{\mathbf{Z}} = \frac{F\angle\theta}{Z\angle\phi} = U\angle\theta - \phi \tag{5-29}$$

and

$$u(t) = \sqrt{2}\, U \cos(\omega t + \theta - \phi) \tag{5-30}$$

In this case we have solved the mechanical circuit by using the techniques developed for electrical circuits. Recognizing that the two circuits are analogous, we could have gone directly to the solution by making appropriate substitutions in Eqs. 5-15 and 5-16. The analogous terms are identified by a comparison of Eqs. 5-25 and 5-26; the results are tabulated in the first two rows of Table 5-1.

Table 5-1 **Analogous Quantities**

Mechanical	M	D	K	f	u	x
Electrical	L	R	C	v	i	q
Rotational	J	D_r	K_r	τ	ω	θ
Thermal		R_T	C_p	τ	q_T	w_T
Electrical II	C	G	L	i	v	λ

The mechanical analog for capacitance C can be obtained by comparing the governing equations for the circuits in Fig. 5.18a and b. Here friction coefficient D (analogous to R) is introduced by the *dash pot* and compliance K is introduced by the spring. From the equations

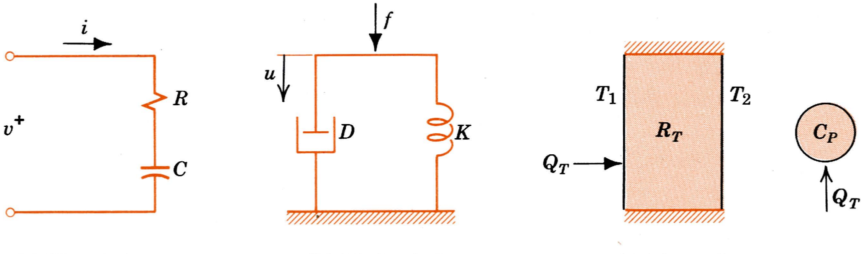
$$v = Ri + \frac{1}{C}\int i\,dt \quad \text{and} \quad f = Du + \frac{1}{K}\int u\,dt \tag{5-31}$$

it is seen that compliance K is the analog of capacitance C. Recalling that displacement $x = \int u\,dt$ and charge $q = \int i\,dt$, we see that x and q are analogous.

Two important equations in heat transfer by conduction are

$$q_T = \frac{\tau}{R_T} \quad \text{and} \quad q_T = C_p\frac{d\tau}{dt} = \frac{dw_T}{dt} \tag{5-32}$$

where q_T = rate of heat flow in cal/s (Fig. 5.18c),
R_T = thermal resistance in s-K/cal,
τ = temperature difference = $T_1 - T_2$ in K,
C_p = thermal capacitance in cal/K,
w_T = thermal energy in cal.



(a) Electrical (b) Mechanical (c) Thermal

Figure 5.18 Other examples of analogous circuits.

The first equation expresses the fact that the rate of heat transfer by conduction is directly proportional to the temperature difference. The second provides for the heat absorbed by a body under conditions of changing temperature. Comparison of these equations with those describing electric circuit components leads to the thermal analogs shown in Table 5-1. Note that in this analog there is no term analogous to mass or inductance; heat flow does not exhibit any momentum or inertia effect.

Equations 5-32 have little practical value because they imply lumped thermal effects, whereas heat conduction is always a distributed phenomenon. (Why?) A further complication is that R_T and C_p vary widely with temperature. In spite of these limitations, writing the equations and identifying the analogous terms are valuable because of the insights gained when the techniques of circuit theory are applied (see Problem 10).

Electrical analogs have been used to study heat flow in power transistors, the production and absorption of neutrons in a nuclear reactor, and the behavior of diverse acoustical and hydraulic systems. In recognition of the ease and effectiveness of the electronic analog computer, the principles and operation of this versatile engineering tool are described in Chapter 19.

Duals

At this time an alert student should be aware of a certain amount of repetition in the discussion of electrical circuits. The equations and diagrams for the series *RL* circuit of Fig. 5.11 are similar to those for the parallel *GC* circuit of Fig. 5.14. With some changes in symbols, the equations and diagrams for the series *RC* circuit of Fig. 5.12 are just like those for the parallel *GL* circuit of Fig. 5.13. Apparently, when we solve one electrical circuit, we automatically obtain the solution to another. How can we take advantage of this interesting fact?

In drawing up the table of mechanical-electrical analogs, we used the series *RLC* circuit (Fig. 5.3). If, instead, we write the integro-differential equation for the parallel *GCL* circuit (Fig. 5.7), we get the different set of analogs shown in the fifth row of Table 5-1.† It is always possible to draw two electrical circuits analogous to a given physical system, and the corresponding terms in the two circuits are related in accordance with the principle of *duality*.

> *When the set of transforms that converts one system into another also converts the second into the first, the systems are said to be duals.*

The series *RLC* circuit and the parallel *GCL* circuit are said to be *duals* because, by using the set of transforms indicated in Table 5-2, either circuit can be converted into the other. In general, the loop equations of a planar network have the same form as the node equations of its dual.† Some of the dual relations

---

† The term λ (lambda) is called flux linkage and is equal to $\int v \, dt$. See Eq. 22-1.

Table 5-2 Dual Quantities

L	R	C	v	i	Z	X
C	G	L	i	v	Y	B

that exist in electrical networks are listed in Table 5-3.

A knowledge of duality enables us to double the benefit from any circuit analysis we perform. As an illustration, the paragraph that includes Eqs. 5-17 and 5-18 transforms directly into the paragraph that includes Eqs. 5-19 and 5-20. To test your understanding of the principle, write out the dual of the discussion of the series *RL* circuit (the paragraph including Eqs. 5-15 and 5-16) and see how well it describes the circuit of Fig. 5.14.

Table 5-3 Dual Relations

Loop current	Node voltage
Kirchhoff's voltage law	Kirchhoff's current law
Series connection	Parallel connection
Current source	Voltage source
Short circuit	Open circuit

Sometimes it is advantageous to construct the dual of a given circuit instead of working with the given circuit itself. The first step (see Fig. 5.19) is to place a dot in each loop of the given circuit and one more dot outside; these dots are the nodes of the dual circuit. Then, through each element of the given circuit, draw a line terminating on the dots. Finally, on each line place the dual of the element through which the line is drawn; these lines are the branches of the dual circuit. The procedure is illustrated in Example 7 on p. 138.

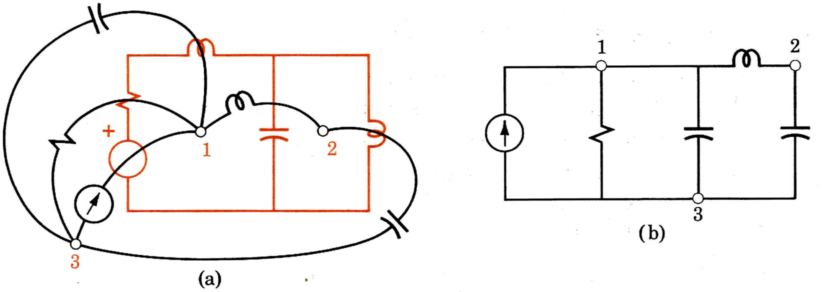


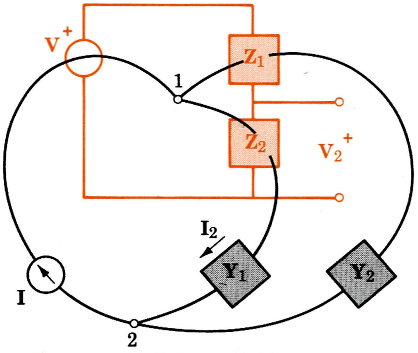
Figure 5.19 Construction of a dual circuit.

---

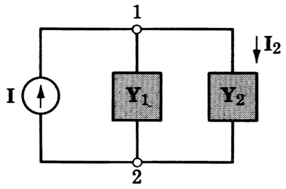
† In the language of network topology, a *planar* network is one which can be drawn on a sphere; every planar network has a dual.

Example 7

The voltage divider of Fig. 5.20a is a useful device. Write an expression for $\mathbf{V}_2$ in terms of $\mathbf{V}$. Then draw the dual circuit and state (without derivation) the dual relation.



(a) Voltage divider



(b) Current divider

Figure 5.20 Deriving the dual of a circuit.

For the *voltage divider,*

$$\frac{\mathbf{V}_2}{\mathbf{V}} = \frac{\mathbf{Z}_2}{\mathbf{Z}_1 + \mathbf{Z}_2} \tag{5-33}$$

The *voltage* across $\mathbf{Z}_2$ is to the total *voltage* as the *impedance* $\mathbf{Z}_2$ is to the sum of the *impedances.*

For the current divider, the dual relation is

$$\frac{\mathbf{I}_2}{\mathbf{I}} = \frac{\mathbf{Y}_2}{\mathbf{Y}_1 + \mathbf{Y}_2} \tag{5-34}$$

The *current* through $\mathbf{Y}_2$ is to the total *current* as the *admittance* $\mathbf{Y}_2$ is to the sum of the *admittances.*

Generality of the Impedance Approach

The impedance approach is a good illustration of the power of a general method. Although the impedance concept was developed for use in determining natural response, it turns out to be applicable to forced response as well. Although it was defined in terms of exponential functions, with a proper interpretation impedance can also be used with direct and sinusoidal currents. Furthermore, when an electrical circuit has been solved, we automatically have the solution to its dual; through the analysis of an electrical circuit, we gain insight into the behavior of other analogous systems. Mastery of this important concept and the associated techniques is well worthwhile.

SUMMARY

- The waveform of the forcing function determines the waveform of the forced response.

- Impedance is defined for exponentials of the form $i = I_0 e^{st}$.
  Impedances are combined in series and parallel, just as resistances are.
  In general, the forced response is governed by $v = Z(s)i$ where

$$Z_R(s) = R \qquad Z_L(s) = sL \qquad Z_C(s) = \frac{1}{sC}$$

- For $s = 0$, $i = I_0 = I$, a direct current. In this case,

$$Z_R(0) = R \qquad Z_L(0) = 0 \qquad Z_C(0) = \infty$$

  An inductance looks like a short circuit to a direct current.
  A capacitance looks like an open circuit to a direct voltage.

- Phasor voltage and current are related by the complex impedance $\mathbf{Z} = Z(j\omega)$ or the complex admittance $\mathbf{Y} = Y(j\omega)$.

$$\begin{array}{llll} \mathbf{V} = \mathbf{ZI} \text{ where } & Z_R(j\omega) = R & Z_L(j\omega) = j\omega L & Z_C(j\omega) = 1/j\omega C \\ \mathbf{I} = \mathbf{YV} \text{ where } & Y_R(j\omega) = G & Y_L(j\omega) = 1/j\omega L & Y_C(j\omega) = j\omega C \end{array}$$

  For phasors, Kirchhoff's laws are written: $\Sigma\mathbf{V} = 0$ and $\Sigma\mathbf{I} = 0$.

- The steady-state sinusoidal response of a two-terminal network is completely defined by $\mathbf{Z}$ or $\mathbf{Y}$ at the terminals where

$$\mathbf{Z} = Z\underline{/\phi_Z} = R + jX \quad \text{and} \quad \mathbf{Y} = Y\underline{/\phi_Y} = G + jB$$

$$\begin{array}{rlcrl} \text{where } \phi_Z &= \tan^{-1}(X/R) & \text{and} & \phi_Y &= \tan^{-1}(B/G) \\ R &= \text{ac resistance in ohms} & & G &= \text{ac conductance in mhos} \\ X &= \text{reactance in ohms} & & B &= \text{susceptance in mhos} \\ &= \omega L \text{ or } -1/\omega C & & &= \omega C \text{ or } -1/\omega L \end{array}$$

- In ac circuits,
  The voltage $\mathbf{V}_R$ is in phase with the current $\mathbf{I}_R$.
  The voltage $\mathbf{V}_L$ leads the current $\mathbf{I}_L$ by 90°.
  The voltage $\mathbf{V}_C$ lags the current $\mathbf{I}_C$ by 90°.

- To determine the forced response to sinusoids:
  1. Transform time functions to phasors and evaluate complex immittances.
  2. Combine immittances in series or parallel to simplify the circuit.
  3. Determine the desired response in phasor form.
  4. Draw a phasor diagram to check values and display results.
  5. Transform phasors to time functions if required.

- Analogous systems are described by similar integro-differential equations. Corresponding terms in the equations are analog quantities. The solution of a problem is applicable to all analogous problems.

- Two networks are duals if the set of transforms that converts the first into the second also converts the second into the first. The loop equations of a planar network have the same form as the node equations of its dual. The solution of any network automatically provides the solution to its dual.

REVIEW QUESTIONS

1. Define impedance.
2. Given $v = V_m \cos \omega t$ and $i = I_m \cos (\omega t + \phi)$, derive an expression for v/i. Is this an "impedance"? Does v/i look useful?
3. Derive the dc element impedances for R, L, and C from the values of $Z(j\omega)$, letting ω approach zero as a limit. In the s plane where $s = \sigma + j\omega$, what is the geometrical interpretation of letting σ approach zero or letting ω approach zero?
4. In terms of steady-state response to a direct current or voltage, what is the effect of an inductance? Of a capacitance?
5. What is meant by a "long time" after closing a switch?
6. What two major advantages result from representing sinusoids by exponentials?
7. Outline in words the steps taken in arriving at the conclusion that phasor voltage and phasor current are related by a complex impedance.
8. Compare the dimensions of $Z_L(j\omega)$ and $Z_C(j\omega)$ with those of ohms.
9. What is the distinction between a phasor and a complex quantity?
10. Outline in words the reasoning that leads to the relation $\Sigma \mathbf{I} = 0$ into any node.
11. Define admittance, reactance, susceptance, and immittance.
12. A two-terminal circuit consists of a capacitance C in parallel with a series combination of resistance R and inductance L. In the admittance $\mathbf{Y} = G + jB$, is $G = 1/R$? What is the value of G?
13. Starting with element immittances on the complex plane, show the graphical determination of $\mathbf{Y}$ and $\mathbf{Z}$ for the circuit in the preceding question.
14. Sketch a sinusoidal voltage wave $v(t)$ and a current $i(t)$ that *leads* the voltage by 45°. Draw the corresponding phasors. If v and i are the inputs to a two-terminal circuit, is it predominantly capacitive or inductive?
15. Given $\mathbf{I} = 10\underline{/150^\circ}$ and $\mathbf{V} = 200\underline{/-150^\circ}$. Is the associated $\mathbf{Z}$ inductive or capacitive? What is the phase angle ϕ?
16. Draw up a table summarizing the voltage-current phase angle relations for inductive and capacitive reactances and susceptances.
17. Define analog and dual.
18. Describe a hydraulic system and draw its electrical analog.
19. Draw a mechanical dual of the *D-M* combination in Fig. 5.17. It may help to write the dual equation first.
20. Draw a phasor diagram showing $\mathbf{U}$ and $\mathbf{F}$ of Eq. 5-29.
21. Why is practical heat conduction "always a distributed phenomenon"?
22. Without reference to the text, list 6 electrical quantities or phrases and their duals.
23. Write out the dual of the discussion of the series RL circuit and see how well it describes the circuit of Fig. 5.14.

EXERCISES

1. In Fig. 5.21, $L = 2\ \mathrm{H}$, $R = 3\ \Omega$, and $C = 0.2\ \mathrm{F}$. If $v_R = 6\,e^{-2t}$ V, predict i_R, v_L, and i.

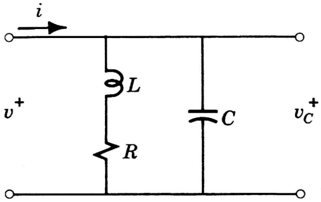


Figure 5.21

2. In Fig. 5.21, $L = 5\ \mathrm{H}$, $R = 20\ \Omega$, and $C = 0.2\ \mathrm{F}$.
   (a) Calculate sufficient points to plot $Y(s)$ for real values of s.
   (b) Determine response current i to voltage $v = 10\,e^{st}$ where $s = 0$, $s = -2$, $s = -4$, and $s = -5$.
   (c) What is the impedance of the circuit to direct current?
3. In Fig. 5.22, $R_1 = 4\ \Omega$, $R = 2\ \Omega$, $L = 2\ \mathrm{H}$, and $C = 0.5\ \mathrm{F}$. Determine the current i a "long" time after applying a direct voltage $v = 12$ V by:
   (a) Calculating $Z(s)$ and letting s approach 0 as a limit;
   (b) Modifying the circuit in accordance with the rules in quotation marks on p. 119, 120.

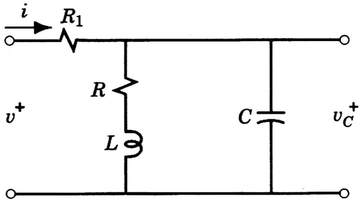


Figure 5.22

4. In Fig. 5.23, determine current I a "long" time after switch S is closed by:
   (a) Calculating $Z(s)$ and letting s approach 0 as a limit;
   (b) Modifying the circuit in accordance with the rules in quotation marks on p. 119, 120.

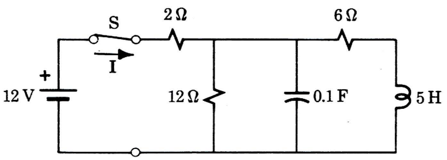


Figure 5.23

5. In Exercise 4, what is the voltage on the capacitance after a "long" time?
6. Given the circuit of Fig. 5.22 where $R_1 = R = 3\ \mathrm{k\Omega}$, $L = 1\ \mathrm{H}$, and $C = 0.25\ \mu\mathrm{F}$. Predict the forced voltage v_C in response to:
   (a) A direct current $i = I = 20$ mA.
   (b) An exponential current $i = 20\,e^{-4000t}$ mA.
   (c) A sinusoidal current $i = 20 \cos(4000t + 60°)$ mA.
7. Consider the circuit of Exercise 1. If $v_R = 6\sqrt{2} \cos 2t$ V, predict phasors $\mathbf{I}_R$, $\mathbf{V}_L$, and $\mathbf{I}$ and show them on a phasor diagram.
8. When a voltage $v = 20 \cos 120t$ V is applied across an impedance, the current is $i = 4 \cos(120t + 37°)$ A; find the real and imaginary components of the complex impedance $\mathbf{Z}$.
9. In Fig. 5.24, $v = V_m \cos \omega t$.
   (a) Using basic element and connection equations (and *no* phasors), derive a literal expression for the steady-state current i in terms of the given quantities only.
   (b) Using the phasor approach outlined in Example 2, derive a literal expression for current i.
   (c) Compare the results of parts (a and b).

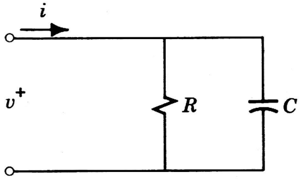


Figure 5.24

10. Given the loop equations:

$$\begin{cases} \mathbf{V}_1 = \mathbf{I}_1(R_1 + jX_C) + j\mathbf{I}_2X_C \\ 0 = j\mathbf{I}_1X_C + \mathbf{I}_2(R_2 + jX_L + jX_C) \end{cases}$$

Draw the network and show the loop currents I_1 and I_2 with proper polarities.

11. A current $i(t) = 2\sqrt{2}\cos(5000t + 30°)$ mA flows in a series combination of $R = 2309\ \Omega$ and $L = 0.8$ H. Determine total voltage $v(t)$:
   (a) By finding the individual phasor voltages and then adding them before transforming $\mathbf{V}$ to $v(t)$
   (b) By finding the total impedance and using $\mathbf{V} = \mathbf{ZI}$.

12. A voltage $v(t) = 20\sqrt{2}\cos(1000t - 50°)$ V is applied across a parallel combination of $R = 2\ \text{k}\Omega$ and $C = 0.5\ \mu\text{F}$. Determine total current $i(t)$:
   (a) By finding the individual phasor currents and then adding them before transforming $\mathbf{I}$ to $i(t)$.
   (b) By finding the total admittance and using $\mathbf{I} = \mathbf{YV}$.

13. In Fig. 5.24, $R = 2\ \text{k}\Omega$ and $C = 2.5\ \mu\text{F}$.
   (a) Determine the two components of a *series* circuit having the same terminal admittance at $\omega = 1000$ rad/s.
   (b) Repeat part (a) for $\omega = 500$ rad/s.

14. A two-terminal network consists of $R = 300\ \Omega$ and $C = 2\ \mu\text{F}$ in series.
   (a) Determine the two components of a *parallel* network having the same terminal impedance at $\omega = 1000$ rad/s.
   (b) Repeat part (a) for $\omega = 2000$ rad/s.

15. In Fig. 5.21 $L = 0.2$ H, $R = 3\ \text{k}\Omega$, and $C = 10$ nF.
   (a) Determine the two components of a series circuit having the same terminal impedance at $\omega = 20$ krad/s.
   (b) Repeat for $\omega = 10$ krad/s.

16. In Fig. 5.25, $R = 20\ \Omega$, $L = 10$ mH, $C = 10\ \mu\text{F}$, and $v = 50\sqrt{2}\cos(2500t + 30°)$ V.
   (a) Calculate individual and total impedances and determine phasors I, V_R, V_L, and V_C.
   (b) Show phasors on a labeled phasor diagram drawn approximately to scale.
   (c) Calculate $i(t)$.

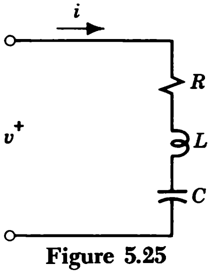


Figure 5.25

17. Assuming the applied voltage v is represented by the phasor $V\underline{/0°}$ construct a labeled phasor diagram showing all voltages and currents in: (a) Fig. 5.24, (b) Fig. 5.25, (c) Fig. 5.21, (d) Fig. 5.22. (Assume relative values of R, X_L, and X_C.)

18. In Fig. 5.21, $X_L = 2\ \Omega$, $R = 2\ \Omega$, $X_C = -2\ \Omega$, $\omega = 1000$ rad/s, and V_L is represented by $\mathbf{V}_L = 20\underline{/90°}$ V.
   (a) Calculate L and C.
   (b) Calculate, in order, and show on a phasor diagram with $\mathbf{V}_L$, the phasors $\mathbf{I}_L$, $\mathbf{V}_C$, $\mathbf{I}_C$, and $\mathbf{I}$.
   (c) Determine $i(t)$.

19. A voltage $v(t) = 10\sqrt{2}\cos(2t + 60°)$ V is applied to the circuit of Exercise 1. Predict $i(t)$.

20. In the circuit of Fig. 5.26, $R = 2\ \Omega$, $L = 3.25$ mH, and $C = 100\ \mu\text{F}$. The steady-state voltage across C is $v_C = 100\sqrt{2}\cos(2000t - 90°)$ V.
   (a) Express $\mathbf{V}_C$ as a phasor and determine $\mathbf{I}_C$ and $\mathbf{V}_L$.
   (b) Using phasor methods, determine $\mathbf{I}_R$ and $\mathbf{I}_{\text{total}}$.
   (c) Show all sinusoidal voltages and currents on a phasor diagram.
   (d) Determine $i(t)$.

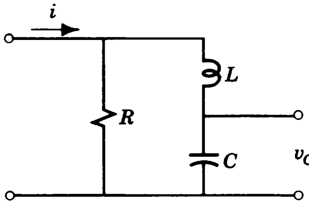


Figure 5.26

21. In Fig. 5.27, the total current is $\mathbf{I} = 10\underline{/0°}$ A and the current through $X_C = -20\ \Omega$ is $\mathbf{I}_C = 4\underline{/127°}$ A. Show these on a phasor diagram. Drawing the phasor diagram as you go along, determine R and reactance X. Is X inductive or capacitive?

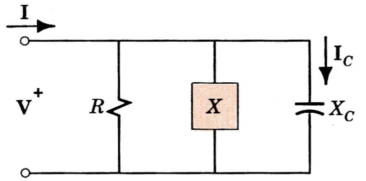


Figure 5.27

22. In Fig. 5.21, $X_C = -50\ \Omega$ and $X_L = 50\ \Omega$. For a certain application, i_L is to lag 120° behind i_C. Specify the value of R.

23. A circuit consists of $R = 35\ \Omega$ in series with an unknown impedance **Z**. For a sinusoidal current with a magnitude of 2 A, observed voltages are 200 V across R and **Z** together and 150 V across the unknown impedance. Find the impedance **Z**.

24. In Fig. 5.28, $L_1 = L = 2$ mH, $C = 500\ \mu$F, and $R = 2\ \Omega$.

(a) For $v = 10 \cos (1000t - 90°)$ V, determine the input impedance.

(b) Calculate currents **I**, $\mathbf{I}_R$, and $\mathbf{I}_L$ and show them on a phasor diagram along with **V**, $\mathbf{V}_{L_1}$, and $\mathbf{V}_C$.

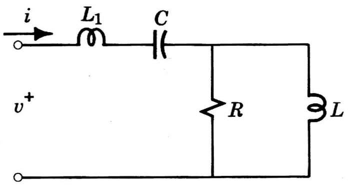


Figure 5.28

25. The load in Fig. 5.29 consists of $G = 0.08$ ℧ in parallel with $B_C = 0.06$ ℧. It is fed by a current source $\mathbf{I} = 45/\underline{0°}$ A and a voltage source $\mathbf{V} = 190/\underline{0°}$ V. If $R = 10\ \Omega$ and $X_L = 4\ \Omega$, predict $\mathbf{I}_1$ and $\mathbf{I}_2$.

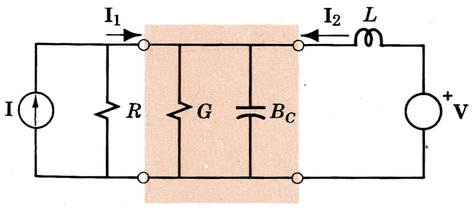
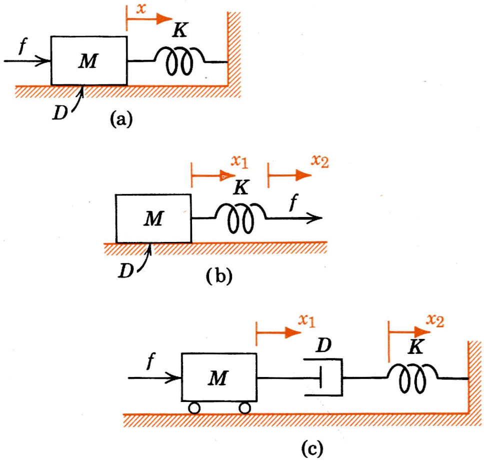


Figure 5.29

26. Write the governing equations for the electrical circuit, the analogous equations for a mechanical (translation) circuit, and draw the mechanical analog for: (a) Fig. 5.24, (b) Fig. 5.25, (c) Fig. 5.21, (d) Fig. 5.26.

27. Write the governing equations for the mechanical circuit, the analogous equations for an electrical circuit, and draw the electrical analog for (a) Fig. 5.30a, (b) Fig. 5.30b, (c) Fig. 5.30c. (Note: $u = dx/dt$ measured from same reference as x.)



(a)

(b)

(c)

Figure 5.30

28. In Fig. 5.30a, $M = 2$ kg, $k = 0.1$ m/N, and $f = 20 \cos 5t$ N.

(a) If $D = 0$, predict $u(t)$ and $x(t)$.

(b) If $D = 6$ N-m/s, predict $u(t)$ and $x(t)$.

29. For the circuit of Fig. 5.31,

(a) Write the governing equations in terms of summation of currents (Σi) into nodes a and b.

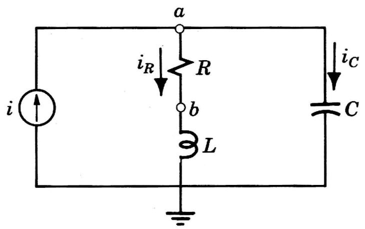


Figure 5.31

(b) Replace each quantity in the two equations by its dual.
(c) Construct the circuit for which the equations of part (b) are the governing equations.
(d) Following the procedure outlined on page 137, construct the dual of the circuit in part (c) and compare with the original.

30. (a) Showing your method, derive the dual of the circuit in Fig. 5.32.
(b) Assume clockwise loop currents i_1 and i_2 and write the governing equations for the given circuit. Write the duals of these equations and compare them with the governing equations for your dual circuit.

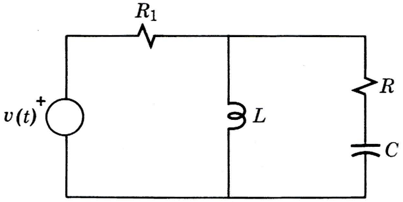
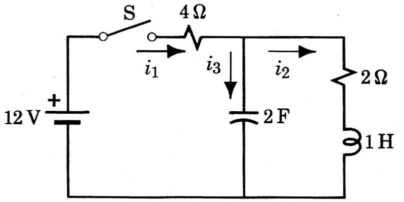


Figure 5.32

31. Construct the dual of Fig. 5.29.

PROBLEMS

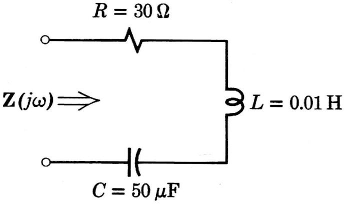
1. Switch S in Fig. 5.33 has been closed a long time. At $t = 0$, the switch is opened. Evaluate i_1, i_2, i_3, v_C, and v_L at $t = 0^-$. Repeat for $t = 0^+$.



S
4 Ω
i_1
i_3
i_2
2 Ω
+
12 V
2 F
1 H

Figure 5.33

2. In Fig. 5.25, $R = 2\,\Omega$, $L = 3$ H, and $C = 0.5$ F. Determine the forced response $i(t)$ for $v(t) = 10t^2$ V where t is in seconds.

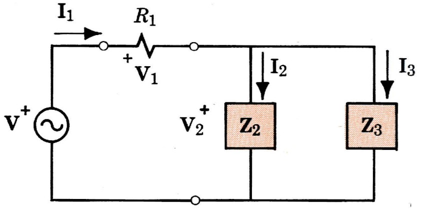


$R = 30\,\Omega$
$\mathbf{Z}(j\omega) \Longrightarrow$
$L = 0.01$ H
$C = 50\,\mu$F

Figure 5.34

3. Given the circuit of Fig. 5.34, design a circuit with only two components that will present the same impedance at $\omega = 1000$ rad/s. Will the two circuits present the same impedance at $\omega = 2000$ rad/s? Explain.

4. In the circuit of Fig. 5.35, voltage $\mathbf{V} = 29.73\underline{/-16.6^\circ}$ V, $\mathbf{V}_2 = 12\underline{/-45^\circ}$ V, $\mathbf{I}_1 = 5\underline{/0^\circ}$ A, and $\mathbf{I}_2 = 3\underline{/-53^\circ}$ A. Design the circuit, i.e., specify values of R_1, $\mathbf{Z}_2$, and $\mathbf{Z}_3$.

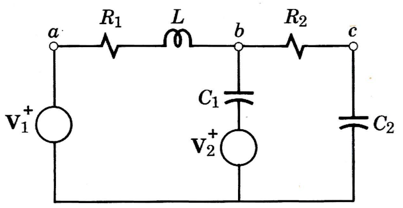


$\mathbf{I}_1$
R_1
+
$\mathbf{V}_1$
$\mathbf{I}_2$
$\mathbf{I}_3$
$\mathbf{V}$ +
+
$\mathbf{V}_2$
$\mathbf{Z}_2$
$\mathbf{Z}_3$

Figure 5.35

5. In Figure 5.36, $\mathbf{V}_1 = 30 + j10$ V, $\mathbf{V}_2 = 30 + j0$ V, $L = 1$ H, $\omega = 1000$ rad/s, $R_1 = R_2 = 1$ kΩ, and $C_1 = C_2 = 1\,\mu$F.
(a) Use loop currents to find $\mathbf{I}_1$ and $\mathbf{I}_2$ through $\mathbf{V}_1$ and $\mathbf{V}_2$.
(b) Use node voltages to find $\mathbf{V}_b$.

6. In Fig. 5.36, voltage $\mathbf{V}_1 = 60 + j20$ V,



a
R_1
L
b
R_2
c
C_1
$\mathbf{V}_1$ +
+
$\mathbf{V}_2$
C_2

Figure 5.36

$\mathbf{V}_2 = 60 + j0$ V, $\omega = 500$ rad/s, $L = 2$ H, $R_1 = R_2 = 1000\ \Omega$, and $C_1 = C_2 = 2\ \mu$F. Find current **I** through C_2 by superposition.

7. In the circuit of Fig. 5.37, $C = 200\ \mu$F, $R_1 = 15\ \Omega$, $R = 4.2\ \Omega$, $L = 5.6$ mH, and $i_1(t) = 7\sqrt{2} \cos 1000t$ A. Specify the necessary applied voltage $v(t)$.

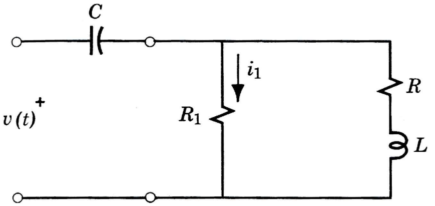


Figure 5.37

8. A 120-V capacitor motor used in a refrigerator can be modeled by the circuit shown in Fig. 5.38. The main winding is $R + jX = 1 + j8\ \Omega$; the auxiliary winding is $R_a + j X_a = 30 + j50\ \Omega$. The mechanical output is represented electrically by $R_o = 10\ \Omega$.
   (a) For good torque production, the current in the auxiliary winding should lead the current in the main winding by 90°. Design the capacitor, i.e., specify the appropriate value of C.
   (b) Assuming the power in R_a and R represents losses, estimate the efficiency of this motor.

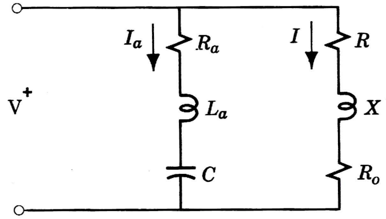


Figure 5.38

9. A "load" of $R = 1000\ \Omega$ is to be supplied by an amplifier designed to provide maximum power to a load of $500\ \Omega$. An experienced engineer suggests the "coupling circuit" of Fig. 5.39. Design the coupling circuit for proper operation at $f = 2000$ Hz.

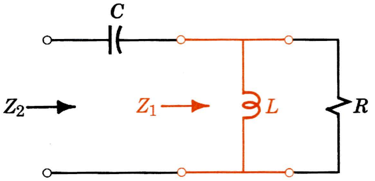


Figure 5.39

10. An unheated house has an effective overall thermal resistance

$$R_T = 0.001°\text{F-hr/Btu}$$

The effective overall thermal capacitance of the contents of the house is $C_p = 6000$ Btu/°F. The outside temperature varies from 100°F at 2 p.m. to 50°F at 2 a.m. Making and stating any necessary assumptions, predict the minimum temperature of the interior and the time at which it will occur.

11. The circuit of Fig. 5.36 has the parameters given in Problem 6. Carefully outlining your reasoning, determine the configuration and the parameters of the dual of Fig. 5.36.

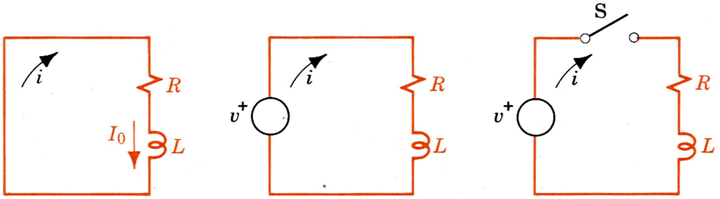
Chapter 6

Complete Response

General Procedure
First-Order Circuits
Second-Order Circuits
Impulse Response

The natural response of a circuit is due to energy stored in inductances or capacitances. Working from the differential equation, we developed a method of using the poles and zeros of the impedance (or admittance) function to indicate the character of the natural response. The forced response is produced by external energy sources such as batteries or generators. Expanding the impedance concept, we then used exponential representation to transform integro-differential equations into algebraic equations, and finding the forced response was reduced to a routine procedure.

In our previous work only one of these responses was considered at a time. In determining natural response (Fig. 6.1a), we assumed energy had been stored by an external source and then that source was removed. In determining forced response (Fig. 6.1b), we assumed that a sufficient time had elapsed so that all natural response components had died away, or at least had become negligibly small. In general, however, there is a transient period during which behavior is not so simple; to determine the current that flows after switch S in Fig. 6.1c is closed we must determine the *complete response*.



(a) Natural (b) Forced (c) Complete

Figure 6.1 Components of circuit response.

Components of the Complete Response

Considering the problem from an energy viewpoint, we note that the natural response tends to decay exponentially as the stored energy is dissipated. The forced response continues indefinitely because energy is supplied to make up any losses. Since we are interested in the transient period in which both effects are present, we conclude that the complete response is some combination of natural and forced responses.

To provide a firm basis for analysis we write the differential equation, using the element equations and the connection equation. Assuming a linear circuit and a sinusoidal forcing function, at all times after closing the switch in Fig. 6.1c,

$$L\frac{di}{dt} + Ri = V_m \cos \omega t \tag{6-1}$$

Any expression that satisfies this equation is a solution of the equation; one such solution is the forced response current

$$i_f = I_m \cos (\omega t + \phi) \tag{6-2}$$

For the related homogeneous equation

$$L\frac{di}{dt} + Ri = 0 \tag{6-3}$$

a solution is the natural response current

$$i_n = I_n e^{-(R/L)t} \tag{6-4}$$

But if the value of i_n from Eq. 6-4 makes the left-hand side of Eq. 6-1 just equal to zero, then the governing equation is satisfied by $i_f + i_n$, or

$$i = i_f + i_n = I_m \cos (\omega t + \phi) + I_n e^{-(R/L)t} \tag{6-5}$$

must also be a solution of Eq. 6-1. We conclude that the complete response is the *sum* of the forced and natural responses.†

Having reached this conclusion, we are in the fortunate position of being able to solve a new group of problems without developing any new theory. The techniques of finding forced and natural response can be applied directly to finding the complete response of a variety of circuits with a variety of driving functions or *excitations*.

In this group are some very practical and important engineering problems. In the design of some systems, the transient behavior is the critical factor; for example, whether the control system of a supersonic aircraft is satisfactory may depend upon its ability to respond smoothly to a sudden disturbance. In other systems, only the transient behavior has any significance; as an illustration, a radar system must generate, transmit, receive, and interpret sharp pulses of energy that never reach a steady state. In this chapter first we develop a general procedure, then we apply it to some first-order circuits with dc and sinusoidal driving functions, and finally we consider the complete response of circuits with more than one energy-storage element.

---

† Mathematicians call these the *particular* solution and the *complementary* solution, respectively. These two components can be added only if the circuit is linear; see Chapter 2, p. 48.

General Procedure

Our general procedure consists of finding the forced and natural response components and then combining them properly. What are the essential characteristics of each component and how are they determined?

Characteristics of the Components

Each component has a distinctive *form* to be discovered and an *amplitude* to be determined. For any excitation that can be described as an exponential, the *form* of the forced response is the same as the form of the forcing function; a direct voltage causes a direct current and a sinusoidal current produces a sinusoidal voltage. The *amplitude* of the forced response is determined by the magnitude of the forcing function and the impedance of the circuit; for direct currents the impedance of interest is $Z(0)$, for sinusoids $Z(j\omega)$.[†]

The *form* of the natural response is governed by the circuit itself. Basically, the form is obtained from the homogeneous differential equation, but our approach is to use the poles and zeros of an immittance function to identify the natural components of voltage and current. The *amplitude* of the natural response is determined from energy considerations. From this viewpoint, the amplitude is just that required to provide for the difference between the actual initial energy storage (in L and C) and that indicated by the forcing function. This idea will become more clear when applied to specific examples.

Procedure

The complete response of any linear two-terminal circuit to exponentials is obtained as follows:

1. *Write the appropriate impedance or admittance function.* The function $Z(s)$ or $Y(s)$ carries all the information of the integro-differential equation based on element and connection equations. Choose the terminals across which the desired voltage appears or into which the desired current flows.
2. *Determine the forced response from the forcing function and the proper immittance.* The form of the forcing function indicates the value of s in $Z(s)$ or $Y(s)$ and the approach used in determining the forced response. For

---

[†] Exponential forcing functions are not common in practical problems and they are not discussed in detail here.

direct currents, apply Ohm's law and $V = RI$. For sinusoids, use phasor methods and $\mathbf{V} = \mathbf{ZI}$.

3. *Identify the natural components from poles or zeros of* $Z(s)$ *or* $Y(s)$. Working with impedance, for example, use the poles to obtain possible components of natural response voltage and the zeros to obtain possible current components. Write these components with undetermined amplitudes.
4. *Add the forced and natural responses and evaluate the undetermined constants.* Obtain the necessary information from initial conditions. From energy considerations we know that inductance currents and capacitance voltages cannot change instantaneously; the energy distribution just after a switch is closed, for example, must be the same as that just before the switch was closed.

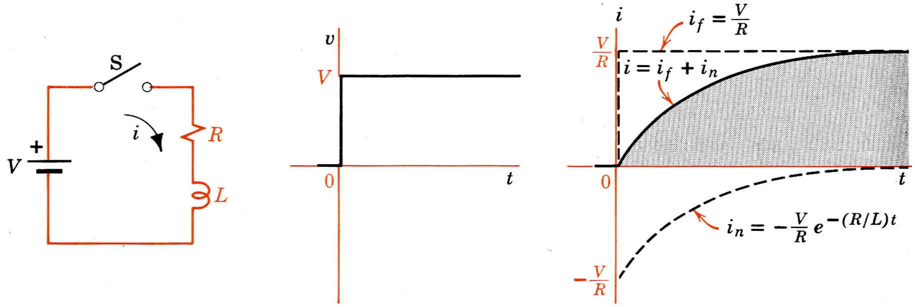
First-Order Circuits

The general procedure is outlined in electrical terms and the emphasis in the following examples is primarily on electric circuits. It should be apparent, however, that the method is applicable to nonelectrical problems as well. The step function that is produced by closing an electrical switch is analogous to suddenly opening a valve, applying a force, or igniting a flame; the complete responses of hydraulic, mechanical, and thermal circuits are similar to those of electric circuits.

Step Response of an RL Circuit

Closing switch S in Fig. 6.2 at time $t = 0$ results in the application of a step voltage of magnitude V to the series combination of resistance R and inductance L. To obtain the complete response current i during the transient period we follow the procedure outlined.

1. *Write the impedance function.*
   For this series circuit, $Z(s) = R + sL$.



(a) (b) (c)

Figure 6.2 Response of an *RL* circuit with a step voltage applied.

2. *Determine the forced response.*
   For a direct voltage source, $s = 0$ and $Z(0) = R$.
   Then $i_f = V/R$ is the forced response.
3. *Identify the natural components.*
   Setting $Z(s) = 0$, $s = -R/L$.
   Therefore, $i_n = A\,e^{-(R/L)t}$ is a possible natural response.
4. *Evaluate the undetermined constants.*
   In general, $i = i_f + i_n = V/R + A\,e^{-(R/L)t}$. Just *before* closing the switch the current in the inductance was zero; therefore, just *after* closing the switch the current must still be zero. (We call this time $t = 0^+$.) At $t = 0^+$

   $$i = I_0 = 0 = \frac{V}{R} + A\,e^{-(R/L)(0)} = \frac{V}{R} + A\,e^0 = \frac{V}{R} + A$$

   Therefore, $A = 0 - V/R = -V/R$, and

   $$i = i_f + i_n = \frac{V}{R} - \frac{V}{R}e^{-(R/L)t} \text{ is the complete response} \qquad (6\text{-}6)$$

Forced, natural, and complete responses are plotted in Fig. 6.2c. The forced response, a current step, has the same form as the forcing function, a voltage step. The natural response is a decaying exponential; the sign is negative because the actual initial energy storage is less than that called for by the forcing function. The complete response current is a continuous function as are all inductance currents. (See Example 1, p. 152.) The initial slope di/dt can be obtained from the differential equation

$$L\frac{di}{dt} + Ri = V$$

by noting that at $t = 0^+$, $i = 0$, $Ri = 0$, and

$$\frac{di}{dt} = \frac{V - Ri}{L} = \frac{V}{L}$$

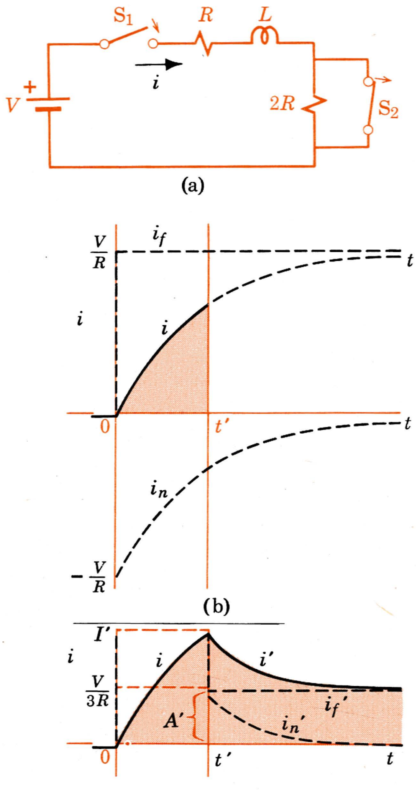
As the current increases, Ri increases and the slope decreases. After several time constants, Ri approaches V and di/dt approaches zero, or the current approaches the steady-state value V/R.

AC Switching Transients

A long time after an ac source is connected to a circuit, the current reaches the steady state with a predictable amplitude and phase relation. In a practical situation, the transient behavior just after the switch is closed cannot be predicted because the natural response depends on the particular point in the cycle at which the switch is closed. Since the "switching transient" may in some cases exceed the rating of the generator and associated transmission equipment, it is

Example 1

In Fig. 6.3a, switch S_1 is open and S_2 is closed. At $t = 0$, switch S_1 is closed and a short time later at $t = t'$ switch S_2 is opened. Determine and plot $i(t)$ for the transient period.



(a)

(b)

(c)

Figure 6.3 Complete response.

Since natural response is initiated at two different times, the problem will be solved in two parts following the four-step procedure.

For $0 < t < t'$,

1. The impedance function is $Z(s) = R + sL$.
2. The forced response is $i_f = V/R$.
3. The natural response is $i_n = A\,e^{-(R/L)t}$.
4. The complete response is $i = i_f + i_n$
$$= V/R + A\,e^{-(R/L)t}$$

At $t = 0^+$, $i = 0 = V/R + A \therefore A = -V/R$ and

$$i = V/R - (V/R)\,e^{-(R/L)t} \qquad (6\text{-}7)$$

as shown in Fig. 6.3b.

For $t' < t < \infty$,

1. The impedance function is $Z'(s) = R + 2R + sL$.
2. The forced response is $i_f' = V/3R$.
3. The natural response is $i_n' = A'\,e^{-(3R/L)(t-t')}$.
4. The complete response is

$$i' = i_f' + i_n' = V/3R + A'\,e^{-(3R/L)(t-t')} \qquad (6\text{-}8)$$

At $t = t'$ just *before* opening switch S_2, by Eq. 6-7

$$i = (V/R)(1 - e^{-(R/L)t'}) = I'$$

At $t = t'^+$ just *after* opening S_2, the current must be the same; by Eq. 6-8,

$$i' = I' = V/3R + A'\,e^0 \therefore A' = I' - V/3R$$

When this value is substituted in Eq. 6-8

$$i' = V/3R + (I' - V/3R)\,e^{-(3R/L)(t-t')} \qquad (6\text{-}9)$$

This response is plotted in Fig. 6.3c, assuming $t' \cong L/R$, one time constant.

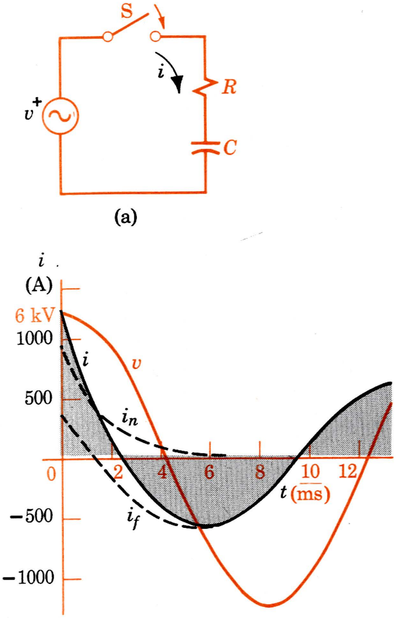
Note that the amplitude of the natural response in each case is determined by the difference between the actual current and the current called for by the forcing function.

essential that the power engineer be able to predict the response under the worst possible conditions.

The method for analyzing switching transients is illustrated in Example 2. Note that the initial current is twice the peak current reached under steady-state conditions. Is this the worst possible condition? Since the initial current is proportional to the initial voltage ($i = v/R$), the largest current will flow if the

Example 2

A load consisting of a series combination of $R = 5\ \Omega$ and $C = 306\ \mu\text{F}$ (Fig. 6.4) is connected to a voltage $v = 6000 \cos 2\pi 60t$ V at time $t = 0$. Find current i as a function of time after closing the switch.



(a)

(b)

Figure 6.4 An ac switching transient.

1. The impedance function is $Z(s) = R + 1/sC$.
2. The forced response is determined for $Z(j\omega)$. Here
$$\begin{aligned}\mathbf{Z} &= R - j1/\omega C\\ &= 5 - j1/377 \times 306 \times 10^{-6}\\ &= 5 - j8.66 = 10\underline{/-60^\circ}\ \Omega\end{aligned}$$
Hence, in terms of maximum value phasors,
$$\mathbf{I}_f = \mathbf{V}_m/\mathbf{Z} = 6000\underline{/0^\circ}/10\underline{/-60^\circ} = 600\underline{/+60^\circ}$$
and
$$i_f = 600 \cos (377t + 60^\circ)$$
3. The natural response is $i_n = A\,e^{-t/RC}$
4. The complete response is $i = i_f + i_n$ as before. However, there is no requirement of continuity of current in a capacitance. Instead, we make use of the fact that voltage on a capacitance cannot suddenly change. Assuming that the capacitance is initially uncharged (the practical case), at $t = 0^+$ the voltage $v_C = 0$ and the full applied voltage appears across R. At $t = 0^+$,
$$\begin{aligned}i = V/R &= 6000 \cos 0^\circ/5 = 1200\\ &= i_f + i_n = 600 \cos 60^\circ + A\,e^0\end{aligned}$$
Therefore, $A = 1200 - 300 = 900$
The complete response is
$$i = 600 \cos (377t + 60^\circ) + 900\,e^{-654t}\ \text{A}$$
as plotted in Fig. 6.4b.

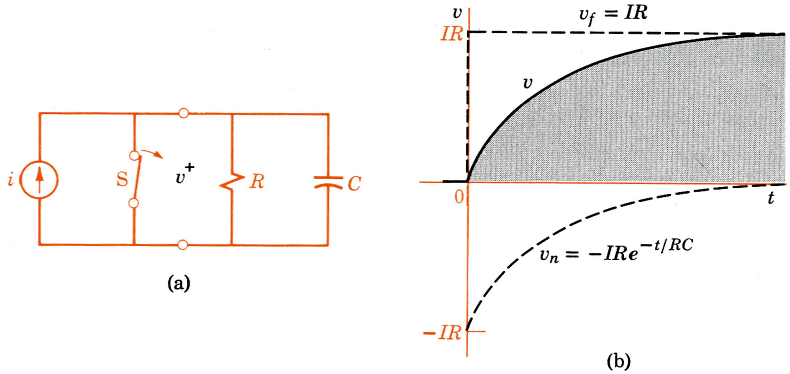
switch is closed at a voltage maximum as in the Example. Is there a point during the cycle at which the switch could be closed with no natural response component? If so, can you find it? (See Exercise 10.)

Series-Parallel Circuits

The step response of the parallel circuit of Fig. 6.5 can be obtained by applying the principle of duality to the result for the circuit of Fig. 6.2. By inspection of Eq. 6-6 and reference to Tables 5-2 and 5-3, we write

$$v = IR - IR\,e^{-t/RC} \tag{6-10}$$

and we have received another dividend from our investment of time and effort in a general principle.



(a)

(b)

Figure 6.5 The step response of a parallel RC circuit.

The circuit of Fig. 6.6 contains another element, and the differential equation approach would be more complicated than for a simple series or parallel circuit. Our impedance approach, however, handles such problems easily, as shown by Example 3. Note the importance of identifying the proper terminals for determining natural response. These are marked in Fig. 6.6a for investigating v.

If i_L were the current of interest, that branch would be opened and the impedance measured "looking in" there with all energy sources "removed." Current sources are removed by open-circuiting and voltage sources by short-circuiting. For determining i_{Ln}, $Z(s) = sL + R$ since open-circuiting I (with switch S open) removes R_1 from consideration. (See Exercise 16.) As an alternative approach, Thévenin's theorem could be used to replace all except L. The resulting equivalent circuit would be a simple series combination of V_T, R_T, and L. The forced and natural components could be identified by inspection or by comparison to Fig. 6.2.

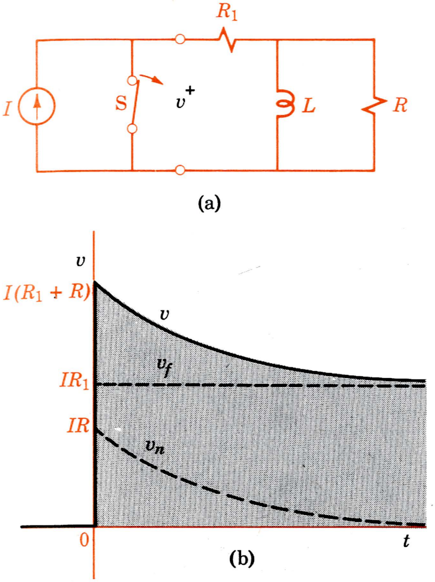
Second-Order Circuits

For a circuit with two different energy-storage elements, the governing differential equation is of the second order. The characteristic equation has two roots or, in other words, the impedance function has two zeros and two poles (one may be at infinity). Working from either interpretation, we conclude that the natural response contains two components and that there are two arbitrary coefficients to be determined. The general procedure for finding the complete response is applicable, although the determination of the constants from initial conditions may be difficult.

Determination of the constants is difficult because each problem is different and no step-by-step procedure can be set down in advance. Where there are two unknown constants, two equations must be written and these are usually for i and di/dt (or v and dv/dt). Useful clues are that i_L and v_C cannot suddenly change and that $v_L = L\,di/dt$ and $i_C = C\,dv/dt$. When these clues are imaginatively combined with Kirchhoff's laws to fit the given circuit, the route to the solution becomes clear.

Example 3

The step current in Fig. 6.6 is obtained from a dc source of amplitude I initially shorted by switch S. At $t = 0$, the switch is opened. Predict the voltage v as a function of time.



(a)

(b)

Figure 6.6 The step response of a series-parallel circuit.

1. The impedance function is

$$Z(s) = R_1 + \frac{R(sL)}{R + sL} = \frac{R_1R + (R_1 + R)sL}{R + sL}$$

2. The forced response is

$$v_f = IZ(0) = I\frac{R_1R}{R} = IR_1$$

3. The natural response voltage is indicated by the pole of the impedance function. $Z(s) = \infty$ for $R + sL = 0$; therefore, $s = -R/L$ and

$$v_n = A\,e^{-(R/L)t}$$

4. The complete response is

$$v = v_f + v_n = IR_1 + A\,e^{-(R/L)t}$$

If switch S has been closed a long time, $i_L = 0$. Just after opening the switch, $i_L = 0$ and all the current flows through R_1 and R in series. At $t = 0^+$,

$$v = I(R_1 + R) = v_f + v_n = IR_1 + A\,e^0$$

Solving, $A = IR$ and

$$v = IR_1 + IR\,e^{-(R/L)t}$$

as shown in Fig. 6.6b.

Step Response of an RLC Series Circuit

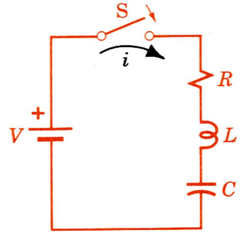


Figure 6.7 An *RLC* series circuit with a voltage step.

As an illustration of the approach to solving second-order circuits, consider the case of a voltage step applied to a series combination of *RLC* (Fig. 6.7). In the typical case, C is initially uncharged and L carries no initial current. The step is obtained from a voltage source V suddenly connected by means of switch S at time $t = 0$. The problem is to find the complete response current i as a function of time. The impedance function is $Z(s) = R + sL + 1/sC$ (step 1) and the forced response is $i_f = V/Z(0) = V/\infty = 0$ (step 2). The natural components (step 3) are obtained by setting $Z(s) = 0$. For

$$s^2L + sR + \frac{1}{C} = 0 \qquad \text{(6-11)}$$

the roots of the characteristic equation are

$$s = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} = -\alpha \pm j\omega \tag{6-12}$$

using the nomenclature of Chapter 4. For roots real and distinct (Eq. 4-12),

$$i_n = A_1 e^{s_1 t} + A_2 e^{s_2 t} \tag{6-13}$$

For roots real and equal (Eq. 4-18),

$$i_n = A_1 e^{st} + A_2 t e^{st} \tag{6-14}$$

For roots complex (Eqs. 4-15, 4-16),

$$i_n = e^{-\alpha t}(B_1 \cos \omega t + B_2 \sin \omega t) \tag{6-15}$$

or

$$i_n = A e^{-\alpha t} \sin (\omega t + \theta) \tag{6-16}$$

For any case, there are two constants to be determined from initial conditions (step 4). Assuming i_L is initially zero, at $t = 0^+$,

$$i = i_f + i_n = 0 \tag{6-17}$$

Assuming v_C is initially zero, at $t = 0^+$, $v_C = 0$ and $v_R = Ri = 0$, and therefore the full applied voltage appears across L or

$$L\frac{di}{dt} = V \quad \text{and} \quad \frac{di}{dt} = \frac{V}{L} \tag{6-18}$$

Equations 6-17 and 6-18 provide two conditions for evaluating the constants.†

For illustration, consider the case of roots real and distinct. From Eq. 6-13, in general

$$\frac{di}{dt} = s_1 A_1 e^{s_1 t} + s_2 A_2 e^{s_2 t}$$

At $t = 0^+$,

$$\frac{di}{dt} = s_1 A_1 + s_2 A_2 = \frac{V}{L} \tag{6-19}$$

and

$$i = i_n = A_1 + A_2 = 0 \tag{6-20}$$

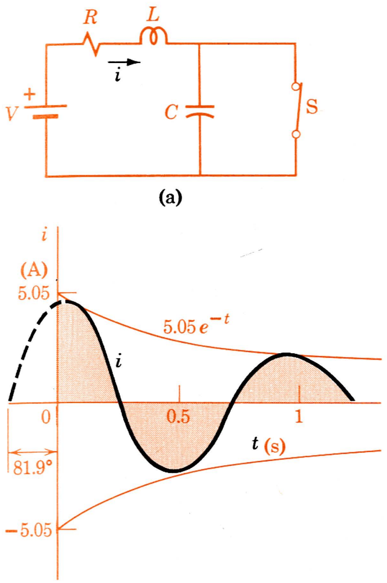
Constants A_1 and A_2 are obtained by solving Eqs. 6-19 and 6-20 simultaneously. The slightly different approach when roots are complex and there is initial energy storage is illustrated in Example 4.

---

†Equation 6-18 can be obtained from the governing equation $V = Ri + L\,di/dt + \int i\,dt/C$ by noting that at $t = 0^+$, $i = 0 \therefore Ri = 0$; also $\int dt = 0 \therefore \int i\,dt/C = 0$. Hence, $V = 0 + L\,di/dt + 0$.

Example 4

In the circuit of Fig. 6.8, $V = 20$ V, $R = 4\,\Omega$, $L = 2$ H, and $C = 0.01$ F. After being closed for a long time, switch S is opened at $t = 0$. Find and plot the complete response current.



(a)

(b)

Figure 6.8 The step response of an *RLC* circuit with energy storage.

Following the general procedure (the steps should be familiar by now),

$$Z(s) = R + sL + 1/sC = 4 + 2s + 100/s$$
$$Z(0) = \infty;\ \text{therefore } i_f = 0$$

Setting $Z(s) = 0$, $s = -1 \pm j7$ are the two roots. For roots complex (Eq. 6-16),

$$i_n = A\,e^{-t}\sin(7t + \theta) = i \qquad (6\text{-}21)$$

At $t = 0^+$, $i = V/R = 20/4 = 5$ and

$$5 = i(0) = A\,e^0 \sin(0 + \theta)\ \therefore\ A\sin\theta = 5$$

As another initial condition, note that $v_C = 0$, and $v_R = iR = V$; therefore, $v_L = 0$ and $di/dt = 0$ at $t = 0^+$. From Eq. 6-21,

$$di/dt = -A\,e^{-t}\sin(7t + \theta) + 7A\,e^{-t}\cos(7t + \theta)$$
$$0 = -A\sin\theta + 7A\cos\theta$$

$\therefore \sin\theta/\cos\theta = 7 = \tan\theta$ or $\theta = \arctan 7 = 81.9°$ and $A = 5/\sin 81.9° = 5.05$.
Therefore, as shown in Fig. 6.8b,

$$i = 5.05\,e^{-t}\sin(7t + 81.9°)\text{ A} \qquad (6\text{-}22)$$

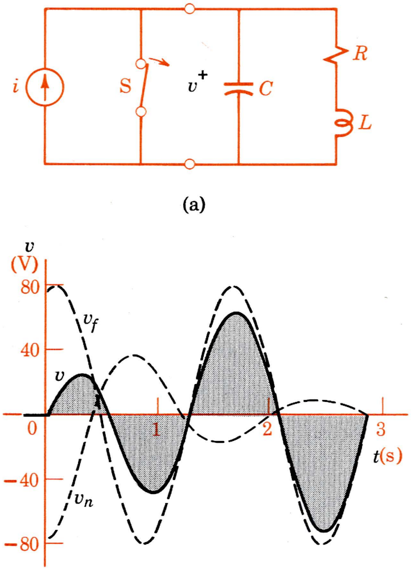
The damping coefficient α indicates a time constant for the envelope of 1 s and the frequency is 7 rad/s or 1.1 Hz.

AC Switching Transients

The sudden application of a sinusoidal forcing function to a series-parallel circuit containing two different energy-storage elements represents the most general case of complete response to be treated in this book. Although no new principles are involved, Example 5 on p. 158 is instructive as an illustration of the techniques employed. Note how $Y(s)$ and $Y(j\omega)$ are used to determine the forced response and the form of the natural response. The amplitude and phase angle of the natural response are determined from energy considerations; at $t = 0$, there is no energy stored in the capacitor ($v_C = 0$) or in the inductor ($i_L = 0$).

Example 5

In the circuit of Fig. 6.9a, a current $i = 8.5 \cos 4t\ A$ is applied to the parallel combination by opening switch S at time $t = 0$. Given that $C = \frac{1}{17}$ F, $R = 2\ \Omega$, and $L = 1$ H, find and plot voltage v as a function of time.



(a)

(b)

Figure 6.9 Buildup of a sinusoidal voltage.

Following the same general procedure,

$$Y(s) = sC + \frac{1}{R + sL} = \frac{s^2LC + sRC + 1}{R + sL}$$

$$Y(j\omega) = \frac{1 - \omega^2LC + j\omega RC}{R + j\omega L} = \frac{1 - \frac{16}{17} + j\frac{8}{17}}{2 + j4}$$

$$= 0.106\underline{/19.5^\circ}\ \mho$$

$$\mathbf{V}_f = \frac{\mathbf{I}}{\mathbf{Y}} = \frac{(8.5/\sqrt{2})\underline{/0^\circ}}{0.106\underline{/19.5^\circ}} = (80/\sqrt{2})\underline{/-19.5^\circ}\ \text{V}$$

For $Y(s) = 0$, $LCs^2 + RCs + 1 = 0$ and

$$s = -R/2L \pm \sqrt{(R/2L)^2 - 1/LC}$$

$$= -1 \pm \sqrt{1 - 17} = -1 \pm j4$$

Therefore, $v_n = A\,e^{-t} \cos(4t + \theta)$, and

$$v_f + v_n = 80 \cos(4t - 19.5^\circ) + A\,e^{-t} \cos(4t + \theta)$$

At $t = 0^+$, $v_C = v = 0$, or

$$0 = 80 \cos(-19.5^\circ) + A \cos\theta \quad \therefore\ A\cos\theta = -75.5$$

Also, $i_L = 0$ $\therefore$ all the current flows in C and

$$\frac{8.5}{C} = \frac{dv}{dt} = -320 \sin(-19.5^\circ) - A\cos\theta - 4A \sin\theta$$

Substituting the value for $A \cos\theta$ and solving,

$$A\sin\theta = 9.4 \text{ and } A\sin\theta/A\cos\theta = -9.4/75.5$$
$$\therefore\ \theta = \arctan(-9.4/75.5) = -7.1^\circ$$

and

$$A = A\cos\theta/\cos\theta = -75.5/\cos(-7.1^\circ) = -76$$

Therefore, the complete response is (Fig. 6.9b)

$$v = 80 \cos(4t - 19.5^\circ) - 76\,e^{-t} \cos(4t - 7.1^\circ)\ \text{V}$$

Example 5 is a special case in that the frequency of the source was selected to be equal to the frequency of the natural response. Under these conditions the voltage oscillations build up from zero and reach the steady-state amplitude in a few cycles. The more general case is illustrated by Exercise 23.

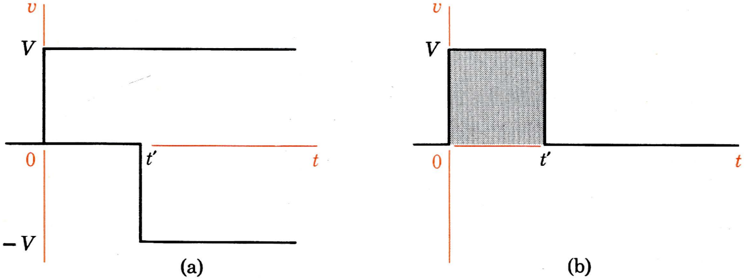
Pulse and Impulse Response

It is appropriate at this point to introduce an abstract concept that is of minor importance in a first course but of great importance in advanced courses. The mechanical analog of this electrical effect is the familiar hammer blow on a mass, giving the mass momentum. Because the analog is well understood, because all the necessary electrical foundation has been laid, and because this discussion rounds out the treatment of complete response it is included here. However, mastery of this concept is not essential to study of the material in this book and this section may be omitted.

Pulse Response

An interesting conclusion from Example 4 is that a step voltage can produce a damped sinusoidal response in an *RLC* circuit. In an analogous way the sudden lateral displacement of a guitar string can produce vibrations that die away exponentially. More commonly the string is rapidly displaced and released or "plucked." The corresponding electrical forcing function is the *pulse*.

As shown in Fig. 6.10, a pulse can be generated by the combination of a positive step and, a short time later, a negative step. This representation is useful because we are familiar with step responses. This is a *rectangular pulse* with an amplitude V and a duration t'.

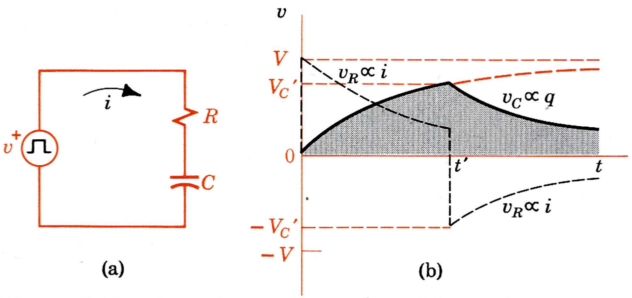


(a) (b)

Figure 6.10 A rectangular pulse formed from two steps.

The response of the *RC* circuit (Fig. 6.11) to such a forcing function is obtained as the response to two voltage steps. There is no forced response because $Z_C(0) = \infty$ and the complete response to the first step is

$$i = i_n = \frac{V}{R} e^{-t/RC} \qquad 0 < t < t' \tag{6-23}$$



(a) (b)

Figure 6.11 The pulse response of an *RC* circuit.

Assuming the capacitor is initially uncharged, the capacitance voltage at time t is

$$v_C = \frac{1}{C}\int_0^t i\,dt = \frac{1}{C}\int_0^t \frac{V}{R}e^{-t/RC}\,dt = \frac{1}{C}\frac{V}{R}(-RC)\,e^{-t/RC}\Big]_0^t$$

$$= -V\,e^{-t/RC}\Big]_0^t = V(1 - e^{-t/T}) \tag{6-24}$$

where $T = RC$ is the time constant.

If a negative step is applied at $t = t'$, the current response has the *form* of Eq. 6-23. Since the capacitance voltage cannot suddenly change, at $t = t'^{+}$, $v_C = V_C'$ as evaluated from Eq. 6-24 and shown in Fig. 6.11b. The current reverses as the capacitance discharges and

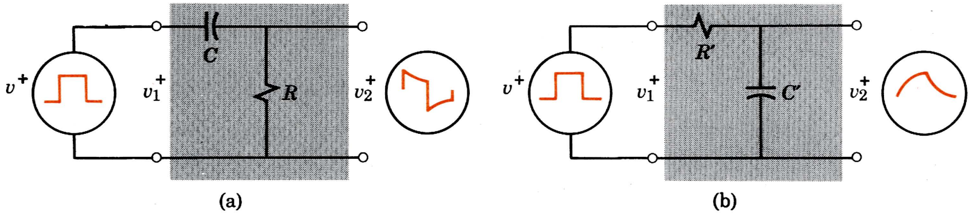
$$i' = i_n' = -\frac{V_C'}{R}e^{-(t-t')/RC} \qquad t' < t < \infty \tag{6-25}$$

Since i is proportional to v_R at all times, a separate current curve is not shown. The charge q on the capacitor is equal to Cv_C and varies as v_C.

Practical Pulses

Pulses are widely used in data processing and in instrumentation. The signals received, processed, and stored in digital computers are usually in pulse form. Rectangular pulses are formed and manipulated by electronic switches operating at very high speeds to provide millions of arithmetic operations per second.

A great deal can be learned about a device by observing its response to short and long pulses. For this purpose, laboratory *pulsers* are available that will generate pulses with duration times of a few nanoseconds to several seconds and repetition rates up to many million pulses per second. To determine the low-frequency characteristics of a device, it might be represented by the model in Fig. 6.12a. When a long pulse is applied at the input, the output ($v_2 = v_R$) will "droop" as in Fig. 6.11b; the amount of droop determines the time constant of the device. (See Exercise 24.) At high frequencies, the same device might be represented by the model in Fig. 6.12b. When a fast-rising pulse is applied at



(a) (b)

Figure 6.12 Device models for use in pulse analysis and typical response.

the input, the output ($v_2 = v_{C'}$) will rise more slowly; to the experienced engineer, the "rise time" indicates the high-frequency characteristics of the device. (See Problem 9).

Impulse Function

The effect of the pulse in Fig. 6.11 is to charge the capacitor, and the pulse response is an exponential discharge current. The amount of charge stored is dependent upon the amplitude of the pulse V and the duration T_P. The product VT_P is the area of the voltage-time curve and is a significant characteristic of the pulse; the same charge can be stored by a shorter pulse with a greater amplitude.

The effect of varying the pulse dimensions while keeping the area constant is shown in Fig. 6.13. If a steady current $I_O = V_O/R$ flowed for a time $T_P = 0.2T = 0.2RC$, the charge stored would be

$$Q_O = I_O T_P = \frac{V_O}{R} T_P = \frac{V_O(0.2RC)}{R} = 0.2CV_O \tag{6-26}$$

Actually, by Eq. 6-24, for the pulse of Fig. 6.13a,

$$Q_a = CV_C = CV_O(1 - e^{-0.2}) = 0.181CV_O = 0.905Q_O$$

For a duration $T_P/2$ and amplitude $2V_O$ (same area),

$$Q_b = CV_C = C2V_O(1 - e^{-0.1}) = 0.19CV_O = 0.95Q_O$$

For a duration $T_P/4$ and amplitude $4V_O$,

$$Q_c = CV_C = C4V_O(1 - e^{-0.05}) = 0.195CV_O = 0.975Q_O$$

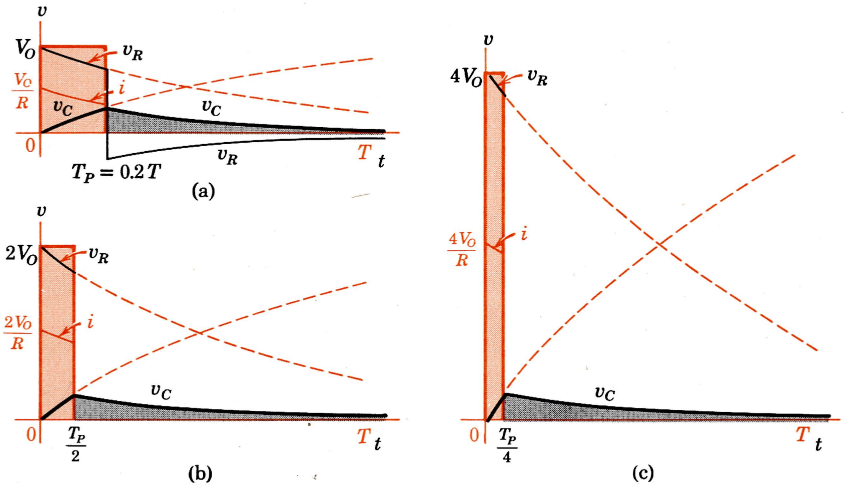


Figure 6.13 The effect of varying the pulse dimensions.

From these calculations we conclude that as the pulse becomes shorter the actual charge stored approaches the charge stored by a rectangular current pulse I_0T_P. The limiting case of a pulse of area I_0T_P is the *impulse,* a pulse of infinite amplitude for an infinitesimal time but with a finite magnitude. The magnitude of a current impulse is $M_P = \int i\,dt$ and the form of $i(t)$ is not important; in fact, in many problems the shape of $i(t)$ is not known, but the response to such an impulse can be determined precisely.

Impulse Response

Three analogous expressions for impulses are:

$$\int f\,dt = MU \qquad \int i\,dt = CV \qquad \int v\,dt = LI \tag{6-27}$$

In each case it is assumed that initially the element is at "rest" with no initial velocity, charge, or current. The effect of a force impulse on a mass is to impart momentum. What is the effect of a current impulse on a capacitance? Since $\int i\,dt = CV$, the effect is to store charge or to establish a voltage on the capacitance. The effect of a voltage impulse on an inductance is to impart *electrokinetic momentum* (LI) or to establish a current. Since there is no forced response to an impulse (Why not?), the complete response is the natural response of the circuit with stored energy. (See Example 6.)

Practical Impulses

The impulse concept can be used to provide initial energy storage without the use of switches; in general network analysis this is advantageous.† Also, in the analysis of ideal circuits impulses occur as part of the complete response; suddenly connecting an ideal current source to an ideal inductor produces an impulse of voltage. Note that the derivative of a step function is an impulse.

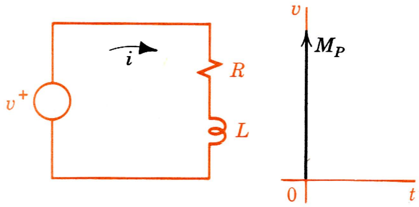
Of more practical significance is the fact that pulses approximating impulses occur in many actual circuits. Sometimes such pulses are unavoidable, as in the case of suddenly applying a voltage generator to a capacitive circuit (see Example 2 in this chapter). Another example is the charge built up on a power transmission line during an electric storm which, when released by a lightning stroke, produces an intense voltage pulse that may destroy the insulation of transformers on the line. In other apparatus, pulses serve valuable functions; a sharp timing pulse is used to trigger the sweep of the electron beam across the screen of a television receiver. If the duration of any pulse is sufficiently short, it can be treated as an impulse by using the approach outlined here.

---

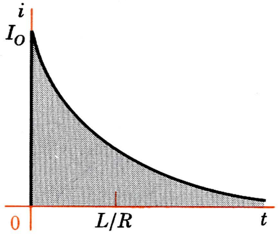
† Another use of impulse functions is in the so-called "operational method" of circuit analysis which is closely related to the Laplace transform method used in more advanced books. See footnote p. 241.

Example 6

The series RL circuit of Fig. 6.14a is "hit" by a voltage impulse of M_P volt-seconds at time $t = 0$. What fraction of the voltage impulse $v(t)$ appears across L? Determine the current response $i(t)$.



(a) Circuit (b) Impulse



(c) Response

Figure 6.14 Impulse response.

The convention for designating an impulse is a vertical arrow drawn at the appropriate instant, as in Fig. 6.14b; the length of the arrow is not significant since the amplitude of all impulses is infinite. (Note the distinction between the *amplitude* of the function and the *magnitude* of the impulse.)

The effect of a voltage impulse is to establish a finite current in L. Since i is finite, $Ri = v_R$ is finite, whereas $v(t)$ is infinite; therefore practically all the voltage impulse appears across L.

By Eq. 6-27, the current established is

$$I = \frac{1}{L}\int v\,dt = \frac{M_P}{L} = I_O$$

The impulse response is then the natural response consisting of the exponential decay of the initial current I_O, or

$$i = i_n = I_O\,e^{-(R/L)t} = \frac{M_P}{L}e^{-(R/L)t} \qquad \text{(6-28)}$$

as shown in Fig. 6.14c.

SUMMARY

- The complete response is the sum of the forced response and the natural response:

$$i = i_f + i_n \qquad \text{or} \qquad v = v_f + v_n$$

- The *form* of the forced response is the same as the form of the forcing function; the *amplitude* is determined by the magnitude of the forcing function and the impedance or admittance.
- The *form* of the natural response is determined by the poles or zeros of the impedance or admittance function; the *amplitude* is determined by the difference between the initial energy storage and that indicated by the forcing function.

- The general procedure for finding complete response is:
  1. Write the appropriate impedance or admittance function.
  2. Determine the forced response from the forcing function and the proper immittance.
  3. Identify the natural components from poles and zeros of $Z(s)$ or $Y(s)$.
  4. Add the forced and natural responses and evaluate the undetermined constants from initial conditions.

- In determining constants from initial conditions, make use of continuity of current in an inductance and voltage across a capacitance. In second-order circuits, also make use of the expressions:

$$\frac{di}{dt} = \frac{v}{L} \qquad \text{and} \qquad \frac{dv}{dt} = \frac{i}{C}$$

- A rectangular pulse can be generated by the combination of a positive step and, a short time later, a negative step.

- An impulse is the limiting case of a pulse of area M_P, amplitude M_P/T_P, and duration T_P as T_P approaches zero.
  For a current impulse, the magnitude $M_P = \int i\,dt = CV$.
  For a voltage impulse, the magnitude $M_P = \int v\,dt = LI$.

- The effect of an impulse is to store energy in an energy-storage element; the impulse response of a circuit is the natural response of the circuit with that initial stored energy.

REVIEW QUESTIONS

1. What is the justification for saying that the complete response is equal to the sum of the forced and natural responses?

2. What determines the form of the forced response? The amplitude?

3. What determines the form of the natural response? The amplitude?

4. How are natural response components related to pole-zero diagrams?

5. Is it possible to determine the natural response components completely before determining the forced response? Explain.

6. Draw a circuit consisting of R and L in parallel, indicate some initial energy storage, and outline the procedure you would follow in determining the complete response to a forcing function.

7. Cite two specific examples of step functions occurring in nonelectrical situations.

8. Write the differential equation for obtaining v in Fig. 6.6 and outline the procedure for finding the natural response. Compare the difficulty of this method with that of solving $Z(s) = 0$.

9. List four commonly employed relations for evaluating undetermined constants from initial conditions.

10. Cite two specific examples of short pulses occurring in nonelectrical situations.

11. What is the distinction between a pulse and an impulse?

12. How can frequency response be deduced from pulse response?

13. In the Summary to Chapter 5, statements are made regarding how L and C "look" to direct currents; formulate similar statements for impulses.

14. Why is there never a "forced response" to an impulse?

15. Why is the actual function $i(t)$ unimportant in determining the response to the impulse $\int i\,dt$?

EXERCISES

1. Demonstrate by substitution that Eq. 6-5 is a solution of Eq. 6-1.
2. (a) Write the governing equation for the circuit of Fig. 6.2a.
   (b) Demonstrate that Eq. 6-6 is a solution of the governing equation.
3. In Fig. 6.15, $v = 20\text{ V}$, $R = 5\text{ k}\Omega$, and $L = 10\text{ mH}$. Predict and sketch current i as a function of time after the switch is closed.

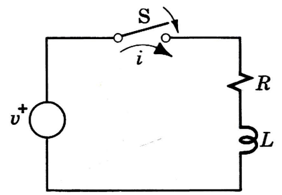


Figure 6.15

4. In Fig. 6.16, a step input is applied to an RC combination at $t = 0$.
   (a) Derive equations for $i(t)$ and $v_C(t)$ if C is initially uncharged.
   (b) For $V = 20\text{ V}$, $R = 2\text{ k}\Omega$, and $C = 5\ \mu\text{F}$, plot $i(t)$ and $v_C(t)$.

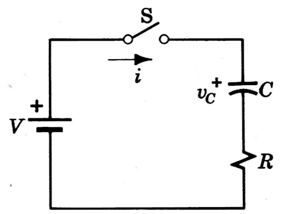


Figure 6.16

5. Repeat Exercise 4 for an initial voltage $v_C = -5\text{ V}$.
6. A battery can be modeled as an ideal source of 20 V in series with a resistance $R_S = 100\ \Omega$. At time $t = 0$, the battery is connected to a series combination of $R = 400\ \Omega$ and $C = 100\ \mu\text{F}$. Predict and sketch the current $i(t)$.
7. In Fig. 6.17, a step input of current is applied to an RL combination by opening switch S at $t = 0$.
   (a) Derive equations for $v(t)$ and $i_L(t)$.
   (b) For $I = 2\text{ A}$, $R = 2\ \Omega$, and $L = 6\text{ H}$, plot $v(t)$ and $i_L(t)$.

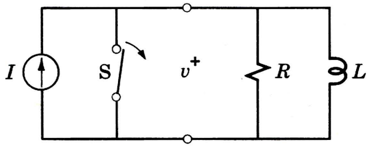


Figure 6.17

8. In Fig. 6.18, S_2 has been closed a long time. At $t = 0$, switch S_1 is closed. For $R_2 = R$:
   (a) Predict and sketch $i(t)$ for $t > 0$.
   (b) Predict and sketch $v_2(t)$.

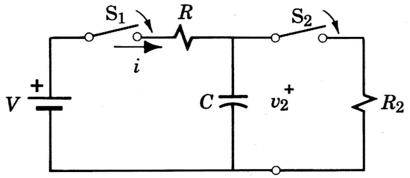


Figure 6.18

9. By means of the circuit in Fig. 6.18, the voltage v_2 is to be brought to 13 V in 3 ms after closing S_1 and held there by closing S_2. If $R = 10\text{ k}\Omega$ and $C = 1\ \mu\text{F}$, "design" the circuit by specifying voltage V and resistance R_2. Sketch $v_2(t)$.
10. Determine the time t at which the switch in Example 2 (Fig. 6.4a) could be closed with no natural response component.
11. In Fig. 6.15, $v = 20 \cos 1000t\text{ V}$, $R = 5\ \Omega$, and $L = 8.66\text{ mH}$. Switch S is closed at $t = 0$.
    (a) Determine $i(t)$ and sketch the components versus time.
    (b) Determine the time t at which the switch could be closed without any natural component of current.
12. (a) At time $t = 0$, switch S in Fig. 6.19 is closed in position 1. Take advantage of Thévenin's theorem to evaluate critical quantities and sketch, approximately to scale on a clearly labeled graph, capacitance voltage $v_C(t)$ for $t > 0$.

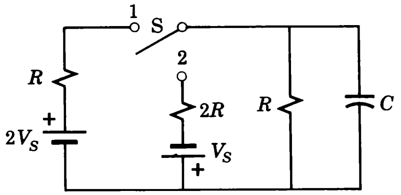


Figure 6.19

(b) At time $t_2 = 3\,RC$, switch S is suddenly thrown to position 2. Sketch $v_C(t)$ for $t > 3\,RC$ on the same graph.

(c) Without detailed derivation, write the equation for $v_C(t)$ after $t = 3\,RC$.

13. In Fig. 6.20, v = a steady voltage V and $R_1 = R$. Following a logical procedure, predict $i(t)$ after closing switch S. Sketch $i(t)$, showing the components.

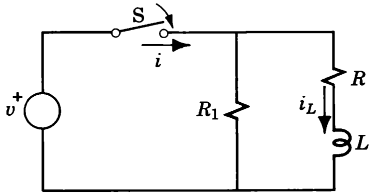


Figure 6.20

14. Repeat Exercise 13, for $v = 20\cos 5t$ V, $R_1 = R = 20\,\Omega$, and $L = 3$ H.

15. Repeat Exercise 13 for $i_L(t)$ in Fig. 6.20. (Hint: Remove v by shorting before evaluating $Z(s)$.)

16. In Fig. 6.21, S_1 is open and S_2 is closed.

(a) Upon closing S_1, what is the initial value of i?

(b) After a long time, what is the value of i?

(c) If S_2 is then opened (S_1 remaining closed), what are the initial values of v_C, i_C, and i_R?

(d) After a long time, what are the values of i and v_C?

17. A step voltage is applied to a series RLC circuit by closing a switch.

(a) Draw the circuit for $V = 10$ V, $R = 150\,\Omega$, $L = 0.5$ H, and $C = 100\,\mu$F.

(b) Predict and sketch the complete current response.

18. Repeat Exercise 17 for $C = 89\,\mu$F, all other values remaining the same.

19. In the series RLC circuit of Fig. 6.7, $V = 70$ V, $R = 4\,\Omega$, $L = 2$ H, and $C = 0.01$ F; there is no initial energy storage. Predict and plot the complete response current $i(t)$.

20. Repeat Exercise 19 for $V = 60$ V, $R = 40\,\Omega$, $L = 4$ H, and $C = 0.01$ F.

21. In Fig. 6.22, $R = 5\,\Omega$, $L = 1.6$ H, and $C = 0.025$ F. Switch S is closed at $t = 0$.

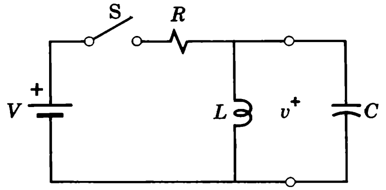


Figure 6.22

(a) Predict *qualitatively*, the voltage $v(t)$ and sketch on the graph.

(b) Predict, *quantitatively*, $v(t)$ for $V = 3$ V.

22. In the circuit of Fig. 6.23, $R = 3\,\Omega$, $L = 3$ H, and $C = 2$ F. The switch is closed at $t = 0$ and then opened at $t = 20$ s.

(a) Sketch $v_C(t)$ for $0 < t < 20$ s.

(b) Sketch $i_L(t)$ for $0 < t < 20$ s.

(c) Estimate the voltage that appears across the switch at the instant it is opened.

(d) What happens to the energy stored in L?

(e) Derive an expression for $v_C(t)$ after the switch is opened.

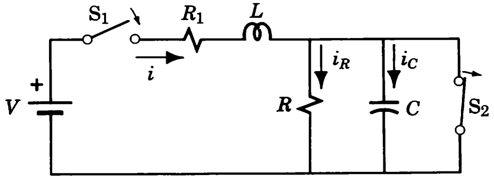


Figure 6.21

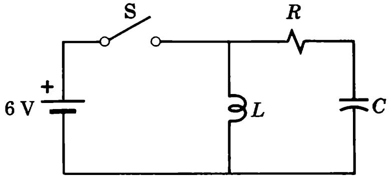


Figure 6.23

23. A voltage $v = 10 \cos 1000t$ V is applied at $t = 0$ to a series *RLC* circuit where $R = 150\ \Omega$, $L = 0.5$ mH, and $C = 100\ \mu$F. Predict and sketch the complete current response.

24. A pulse of 1 ms duration is applied to a device represented by the model of Fig. 6.12a. As observed on an oscilloscope, the output signal "droops" 20%; i.e., the voltage at the end of the pulse is only 80% of the initial magnitude (see Fig. 6.11b). Sketch the input and output waveforms and determine the time constant of the model.

25. A series combination of $R = 5\ \Omega$ and $L = 20$ H is "hit" by an impulse of 40 V-s. Determine the current response $i(t)$.

26. A parallel combination of $R = 200\ \Omega$ and $C = 20\ \mu$F is "hit" by an impulse of 100 A-s. Before and after the impulse, the current through the terminals is zero or the terminals may be considered open. Determine:
(a) The charge carried by the impulse.
(b) The voltage distribution at $t = 0^+$.
(c) The current response in the parallel circuit.

27. In Fig. 6.24, $L = 1$ H and $C = 1\ \mu$F; there is no initial energy storage. At $t = 0$, a rectangular pulse of amplitude 500 V and duration 10 μs is applied. Stating any simplifying assumptions, determine the current just after $t = 10\ \mu$s and the current $i(t)$.

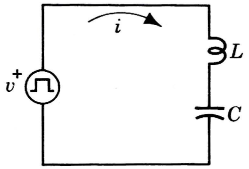


Figure 6.24

PROBLEMS

1. You need a light that flashes automatically 2 times per second. Available is a gaseous discharge lamp that is nonconducting until the applied voltage reaches 30 V and then is highly conducting. Also available are a 60-V battery and a 10-μF capacitor. Devise a simple *RC* circuit that will provide automatic flashing and specify the appropriate value of *R*.

2. In the circuit of Fig. 6.25, switch S_1 is closed at $t = 0$.
(a) Sketch $i(t)$ and determine the time t_1 required for the current i to reach 1.0 A.
(b) S_2 is then closed at t_1; sketch on the same graph $i(t)$ after t_1.
(c) Sketch on the same graph, voltage $v_L(t)$ before and after S_2 is closed.

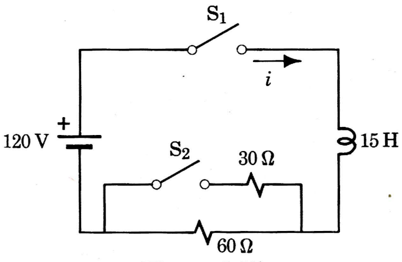


Figure 6.25

3. An ac generator can be represented by a series combination of $v = 2000 \cos 500t$ V and $R = 5\ \Omega$. At time $t = 0$, it is connected across $R = 40\ \Omega$ and $C = 50\ \mu$F in *parallel.*
(a) Draw and label the circuit diagram.
(b) Determine and graph generator current $i(t)$, showing components.
(c) Compare the peak current during the transient period with the maximum reached during steady-state operation.

4. In Fig. 6.26, switch S has been open for a long time; at time $t = 0$, switch S is closed. You are to predict the current $i(t)$.
(a) Draw the pole-zero diagram of imped-

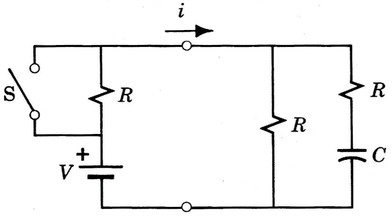


Figure 6.26

ance for the right-hand portion of the circuit.

(b) Derive an expression for $i(t)$, following a logical, step-by-step procedure.

(c) Sketch $i(t)$ during the transient period. Indicate your time scale.

5. In Fig. 6.26, capacitance C is replaced by inductance L. Switch S has been open a long time; at time $t = 0$, S is closed. Derive an expression for the current $i(t)$, and sketch $i(t)$ during the transient period.

6. In Fig. 6.27, switch S has been in position 1 a long time, applying generator voltage $v = 5 \cos 1000t$ V to the series circuit consisting of $R = 30\,\Omega$ and $L = 40$ mH. To achieve a certain desired effect, at an instant when the generator voltage is a maximum (at $t = 0$, for example) the switch is thrown to position 2 and $V = 5$ V dc is applied. Predict and sketch $i(t)$ before and after the switch is thrown.

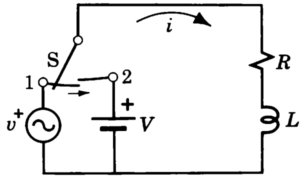


Figure 6.27

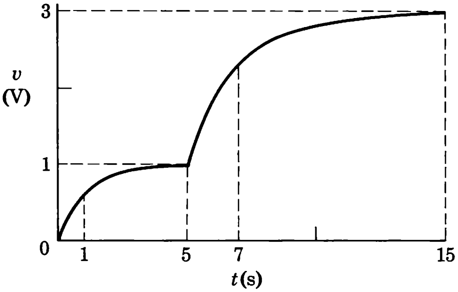


Figure 6.28

7. When a two-position (single-pole, double-throw) switch is thrown to position 1 at $t = 0$ and to position 2 at $t = 5$ s, the voltage across a 200-μF capacitor is to vary with time as shown in Fig. 6.28. Using a single battery, design a circuit to provide the specified voltage variation.

8. The secondary voltage in an automobile ignition coil is given by $v_2 = M\, di/dt$ where i is the primary current (Fig. 6.29). In a certain coil $L = 20$ mH, $R = 50\,\Omega$ and $M = 5$ H. Determine the maximum value of v_2 on the "make" (when S is closed). With a "condenser" $C = 0.2\,\mu$F across the "breaker points," determine the maximum value of v_2 on the "break." State any simplifying assumptions.

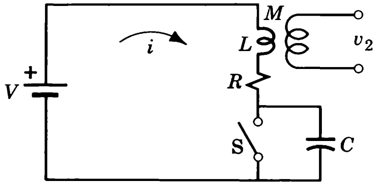


Figure 6.29

9. The high-frequency response of a device (represented by the model of Fig. 6.12b) is to be inferred from the pulse response. When a 10-μs pulse is applied to the input, the rise time of the output is observed on an oscilloscope to be 2 μs. Sketch the input and output waveforms and determine the time constant of the model. Predict the frequency at which the amplitude of the output voltage V_2 will drop to $1/\sqrt{2}$ times that of the input voltage V_1.

10. The door chime of Fig. 23.19a is hit by a plunger. Draw a simplified mechanical model of the chime (using lumped elements) and the analogous electrical circuit. Sketch the vertical velocity of the center of the chime as a function of time.

Chapter 7

Steady-State AC Circuits

Power Calculations
Frequency Response
Resonance
Three-Phase Circuits

The forced response to a periodic function is a steady-state behavior even though voltages and currents are continuously varying. The great bulk of electrical power is generated, transmitted, and utilized in the form of steady-state alternating currents. Also, much of the communication of information by wire or by radio is in the form of ac signals ranging in frequency from a few cycles per second to billions of cycles per second. Certainly the most common, but not the most difficult, problems encountered in electrical engineering involve the analysis of ac circuits, and many books have been written on this important subject.

These are practical problems arising in the design and application of electrical apparatus. They are of concern to the astronautical engineer planning a satellite transmitter installation, the chemical engineer selecting remote process controls, the construction engineer planning a temporary power line, the mechanical engineer building a vibration amplifier, and the industrial engineer improving plant efficiency.

Previous chapters in this book provide the necessary foundation of basic laws, phasor representation, complex immittance, and circuit analysis. Building on that foundation, we now develop techniques and approaches that transform difficult problems into routine exercises. First we study the variation of power in circuit elements and learn to make practical power calculations. Then we consider circuits in which the frequency of the forcing function is varied and study the interesting phenomenon of resonance. Finally we acknowledge the fact that most of the billions of kilowatt-hours of electrical energy used each year is in polyphase form and take a quick look at three-phase circuits.

Power Calculations

Since voltage is the energy per unit charge and current is the charge per unit time, the basic expression for electrical energy per unit time or *power* is

$$p = vi = i^2R = \frac{v^2}{R} \tag{7-1}$$

in a resistance. For sinusoidal variation of voltage and current, the instantaneous value of power is a periodic function. Because average power P is usually the important quantity, we have defined the effective or rms current so that power in a resistance is given by

$$P = \frac{1}{T}\int_t^{t+T} i^2R\,dt = I^2R = \frac{V^2}{R} \tag{7-2}$$

where $I = I_{\text{rms}}$ and $V = V_{\text{rms}}$. This is the most commonly used value of current (or voltage) and is written without a subscript. Other values are defined in terms of the effective value; for example, a "110-volt 60-cycle" house circuit that actually *measures* 115 V rms on an ac voltmeter supplies a voltage $v = \sqrt{2}\,115 \cos 2\pi 60t$ V.

Power is measured with a *wattmeter*. In a common type of wattmeter, the voltage v applied to the "voltage coil" establishes a magnetic field strength directly proportional to v. The current i flowing in the "current coil" reacts with the magnetic field to produce a torque proportional to the instantaneous vi product. The deflection of the needle is proportional to the developed torque and the inertia of the meter movement performs the desired averaging operation. The resulting deflection is proportional to the average vi product and the scale is calibrated to read average power in watts. For sinusoidal voltages and currents applied to the proper coils, the wattmeter reading is equal to $VI\cos\theta$ (Eq. 7-7).

Reactive Power

Since power in a resistance is proportional to the square of current or voltage, the power is always positive and energy is dissipated throughout the entire cycle. In contrast, inductance and capacitance store but do not dissipate energy. When current through an inductance is increasing, energy ($w_L = \frac{1}{2}Li^2$) is transferred from the circuit to the magnetic field; but when the current decreases this energy is returned. Similarly, when the voltage across a capacitance is increasing, energy ($w_C = \frac{1}{2}Cv^2$) is transferred from the circuit to the electric field and power is *positive*. When the voltage decreases, this energy is returned and power is *negative*.

From the basic element relations, for an inductance L carrying a current

$$i = I_m \cos(\omega t - \pi/2) = \sqrt{2}I \sin \omega t$$

the voltage is

$$v_L = L\,di/dt = \sqrt{2}\omega LI \cos \omega t$$

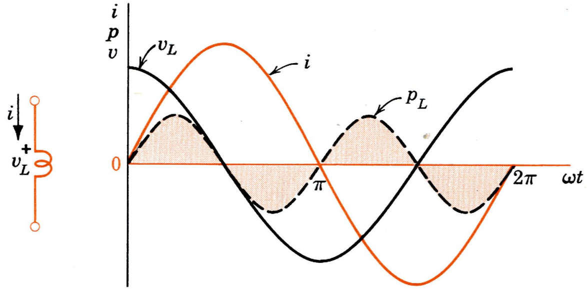


Figure 7.1 The variation of power in an inductance.

and the instantaneous power (Fig. 7.1) is

$$p_L = iv_L = 2I^2\omega L \sin \omega t \cos \omega t = I^2 X_L \sin 2\omega t \tag{7-3}$$

where $X_L = \omega L$ is the inductive reactance. The amplitude of the power variation is I^2X_L; the sin $2\omega t$ factor indicates the periodic return of energy to the circuit.

For a capacitance C carrying the same current, the voltage is

$$v_C = (1/C) \int i\, dt + V_0 = -\sqrt{2}(1/\omega C) I \cos \omega t$$

and the instantaneous power (Fig. 7.2) is

$$p_C = iv_C = 2I^2(-1/\omega C) \sin \omega t (\cos \omega t) = I^2 X_C \sin 2\omega t \tag{7-4}$$

where $X_C = -1/\omega C$ is the capacitive reactance. We conclude that in a reactive element there is no net energy transfer or the average power is zero. However, there is a periodic storage and return of energy and the amplitude of the power variation is

$$P_X = I^2 X \tag{7-5}$$

This useful quantity is called *reactive power* because of its similarity to *active power* $P = I^2R$. Reactive power can also be expressed as V^2/X.

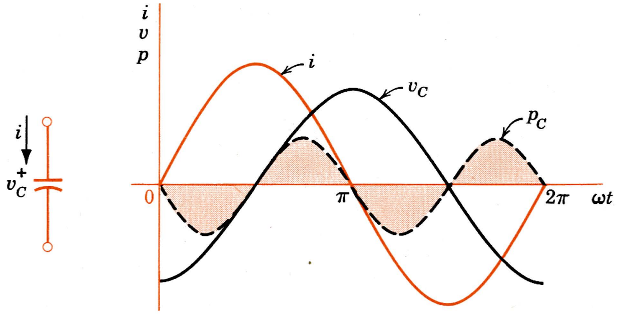


Figure 7.2 The variation of power in a capacitance.

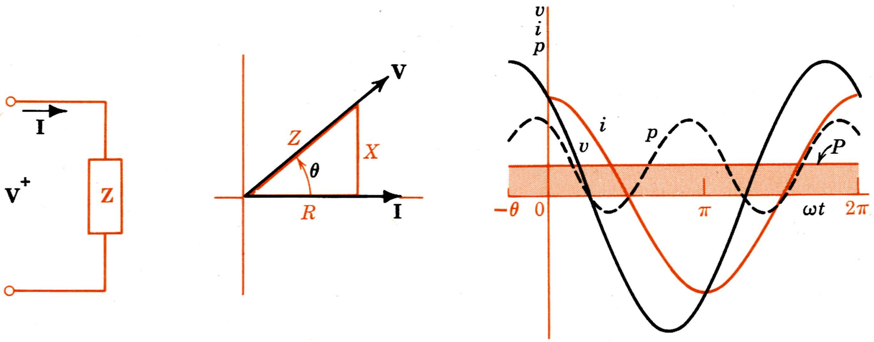
Power Factor

In the general case of alternating current supplied to a complex impedance (Fig. 7.3), voltage and current differ in phase by an angle θ. In phasor notation, $\mathbf{V} = \mathbf{ZI} = (Z/\underline{\theta})\mathbf{I}$; in terms of effective values, $V = ZI$. The instantaneous power is

$$p = vi = \sqrt{2}V \cos(\omega t + \theta) \cdot \sqrt{2}I \cos \omega t = 2VI \cos(\omega t + \theta) \cos \omega t$$

Recalling that $2 \cos A \cos B = \cos(A - B) + \cos(A + B)$, we write

$$p = VI \cos \theta + VI \cos(2\omega t + \theta) \tag{7-6}$$



(a) Circuit diagram (b) Phasor diagram (c) Time variation

Figure 7.3 Power in a general ac circuit.

Integrating over a cycle, or noting that the term $VI \cos(2\omega t + \theta)$ equals zero over a cycle, we find that the average power is

$$P = VI \cos \theta \tag{7-7}$$

In a circuit containing resistance only, voltage and current are in phase, $\theta = 0$, and $P = VI$ in watts† where V and I are effective values in volts and amperes, respectively. In general, the average power P is VI multiplied by a factor that can never exceed unity. By definition, the *power factor* is given by

$$\text{pf} = \cos \theta = \frac{P}{VI} \tag{7-8}$$

In a typical distribution system, components are connected in parallel across a common voltage. An inductive component in which the current lags the voltage (as in Fig. 7.3) is said to have a *lagging power factor;* a capacitive component is said to have a *leading power factor* because the current leads the voltage.

---

†To quote an early power engineer: "We didn't know what watts were; volt-amperes were a measure of power. If anything happened so that the power did not get into the volt-amperes—well, that was the manufacturer's fault. Nobody raised questions, and nobody answered any." P. N. Nunn, *General Electric Review,* September 1956, p. 43.

Since $X = Z \sin\theta$ (reactance X is the imaginary component of the impedance) and $V = ZI$, Eq. 7-5 becomes

$$P_X = I^2 X = I^2 Z \sin\theta = VI \sin\theta \tag{7-9}$$

The quantity $\sin\theta$ is sometimes called the *reactive factor* because of its resemblance to the power factor. The units of reactive power P_X are *volt-amperes reactive* (abbreviated VAR) to emphasize the difference between this quantity and the power in watts.

Example 1

A coil is to be represented by a linear model consisting of inductance L in series with resistance R. In the laboratory, when a 60-Hz current of 2 A (rms) is supplied to the coil (Fig. 7.4), the voltmeter across the coil reads 26 V (rms). A wattmeter indicates 20 W delivered to the coil. Determine L and R.

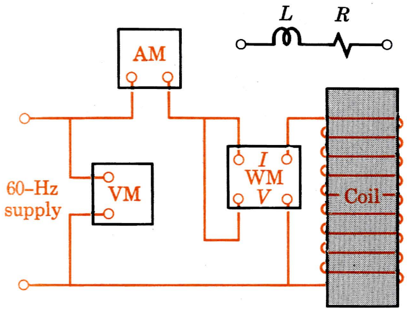


Figure 7.4 Power measurement.

From the basic expression for power (Eq. 7-2),

$$R = \frac{P}{I^2} = \frac{20}{2^2} = 5\ \Omega$$

From the definition of power factor (Eq. 7-8),

$$\theta = \cos^{-1}\frac{P}{VI} = \cos^{-1}\frac{20}{2 \times 26} = 67.4^\circ$$

From the definition of reactive factor (Eq. 7-9),

$$P_X = VI \sin\theta = 26 \times 2 \sin 67.4^\circ = 48\ \text{VAR}$$

From the expression for reactive power (Eq. 7-5),

$$X_L = \frac{P_X}{I^2} = \frac{48}{2^2} = 12\ \Omega$$

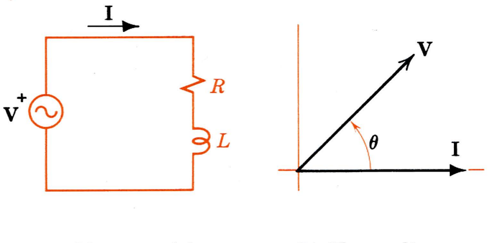
$$L = \frac{X_L}{\omega} = \frac{X_L}{2\pi f} = \frac{12}{2\pi 60} = 0.0318\ \text{H}$$

Check:

$$\begin{aligned} V = ZI &= \sqrt{R^2 + (\omega L)^2}\, I \\ &= \sqrt{5^2 + (377 \times 0.0318)^2} \cdot 2 = 13 \times 2 = 26\ \text{V} \end{aligned}$$

Complex Power

The definition of ac power $VI \cos\theta$ and reactive power $VI \sin\theta$ provides the basis for a new abstract concept that is useful in solving industrial problems in power generation and consumption. Consider an induction motor carrying a current I and represented by the linear model of Fig. 7.5a. The phasor diagram,



(a) Linear model (b) Phasor diagram

Figure 7.5 Representation of an industrial load.

using effective values, shows current **I** lagging voltage **V** by an angle θ in this inductive load. The phasor diagram is repeated in Fig. 7.6b, which also shows the real and imaginary components of the supply voltage. If we *divide* the magnitude of each voltage phasor by I, we obtain a similar triangle, the *impedance triangle* (Fig. 7.6a). If we *multiply* each voltage magnitude by I, we obtain another similar triangle, the *power triangle* (Fig. 7.6c).

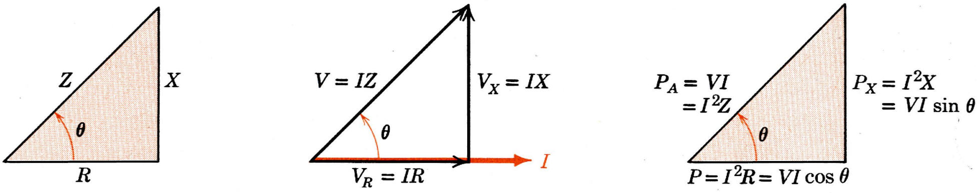
The hypotenuse of the power triangle is a new quantity *apparent power* defined by

$$\mathbf{P}_A = VI\cos\theta + jVI\sin\theta = VI\underline{/\theta} \tag{7-10}$$

or by

$$\mathbf{P}_A = P + jP_X = P_A\underline{/\theta} \tag{7-11}$$

Apparent power is measured in *volt-amperes* or *kilovolt-amperes* (abbreviated VA or kVA). In Fig. 7.6c we can visualize apparent power in VA as a complex quantity with a real component equal to power in watts and an imaginary component equal to reactive power† in VAR.



(a) Impedance triangle (b) Phasor diagram (c) Power triangle

Figure 7.6 Derivation of the complex power triangle.

In deriving Eq. 7-8 for power factor and Eq. 7-9 for reactive power, we assumed that "voltage and current differ in phase by an angle θ." To eliminate an ambiguity, we now specify that *θ is the angle associated with the equivalent impedance.* When the current lags the voltage, θ is positive and therefore reactive power $VI\sin\theta$ is positive. (The power $VI\cos\theta$ is always positive if $|\theta| \le 90°$.) With this interpretation, lagging pf reactive power is positive and leading pf reactive power is negative. In Example 1, the meter readings do not indicate whether θ is positive or negative; our convention indicates that θ is positive and P_X is identified as a positive 48 VAR.

---

† Instead of P_X, the conventional symbol for reactive power is Q; but already we use this symbol for charge and in the next section we shall use Q for another important quantity. To reduce (but not eliminate) confusion, we shall be unconventional here.

The apparent power is a practical measure of the capacity of ac equipment. In a transformer used to "step down" the voltage from a distribution potential of 4000 V to a safe handling potential of 120 V, the allowable output is limited by transformer heating due to losses. The losses (discussed in detail in Chapter 22) are determined by voltage and current and are unaffected by power factor. The size of the transformer, for example, required to supply a given industrial load is determined by the apparent power in volt-amperes instead of by the power in watts.

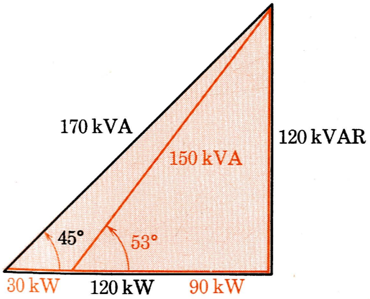
The concept of complex power provides another approach to solving steady-state ac problems. Calculations follow the rules of complex algebra and vector techniques and graphical methods are applicable. Powers in different portions of a network can be added directly, and reactive powers can be added with proper attention to sign. Drawing a power triangle is a valuable aid in visualization. The approach is illustrated in Example 2.

Example 2

An industrial load consists of the following: 30 kW P of heating and 150 kVA (input) P_A of induction motors operating at 0.6 lagging pf. Power is supplied to the plant at 4000 V. Determine the total current and plant power factor.

Load	P_A (kVA)	θ (°)	pf	P (kW)	P_X (kVAR)
Heating	30	0	1.0	30	0
Motors	150	53	0.6	90	120
Plant	170	45	0.71	120	120

(a) Table of power quantities



(b) Power triangles

Figure 7.7 Power factor calculation.

In general, apparent power $\mathbf{P}_A$ is given by

$$P + jP_X = P_A\underline{/\theta}$$

Note that if any two of the four quantities are known, the other two can be calculated. To keep track of the quantities involved, set up the table of Fig. 7.7a where the given quantities are in color.

Assuming that the pf of the heating load is unity, $\theta_H = 0$ and

$$P_A = \frac{P_H}{\cos\theta_H} = \frac{30}{1.0} = 30\text{ kVA}$$

The power "triangle" for this load is the horizontal line segment labeled 30 kW in Fig. 7.7b.

For the motor, $\theta = \cos^{-1} 0.6 = 53°$ and the power components are:

$$P_M = P_A \cos\theta_M = 150 \cos 53° = 90\text{ kW}$$
$$P_{XM} = P_A \sin\theta_M = 150 \sin 53° = 120\text{ kVAR}$$

From the power triangle, for the plant

$$P = P_H + P_M = 30 + 90 = 120\text{ kW}$$
$$P_X = P_{XH} + P_{XM} = 0 + 120 = 120\text{ kVAR}$$

Then

$$\mathbf{P}_A = 120 + j120 = 170\underline{/45°}\text{ kVA}$$

The plant power factor is $\cos\theta = 0.71$. The plant current is

$$\frac{P_A}{V} = \frac{170\text{ kVA}}{4\text{ kV}} = 42.5\text{ A}$$

Power Factor Correction

To supply the plant in Example 2, the local power company must install a transformer capable of supplying 170 kVA (neglecting provision for overloads) and wire large enough to carry 42.5 A. In addition to the 120 kW of power supplied to the plant, the power company must generate power to supply the I^2R losses in all the generation, transmission, and distribution equipment.

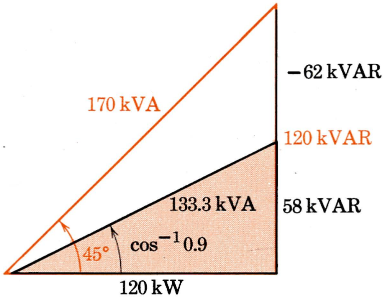
To supply the same power to another plant operating at unity pf would require a transformer rated at only 120 kVA and wire rated at $120/4 = 30$ A, and the I^2R losses in all equipment would be smaller. It is reasonable to expect that the charge for supplying power to this second plant would be lower, reflecting the lower cost to the power company.

Example 3

The power factor of the plant in Example 2 is to be corrected to 0.9 lagging. Specify the necessary auxiliary equipment.

Load	P_A (kVA)	θ (°)	pf	P (kW)	P_X (kVAR)
Previous	170	45	0.71	120	120
Desired	133.3	25.8	0.9	120	58
Correction	62	90	0.0	0	−62

(a) Table of power quantities



(b) Power triangles

Figure 7.8 Power factor correction.

The power of 120 kW is unaffected by adding reactive power equipment. A table of given values (in color) and calculated values is shown in Fig. 7.8. The magnitude of the apparent power desired after correction is

$$P_A = \frac{P}{\cos\theta} = \frac{120}{0.9} = 133.3 \text{ kVA}$$

and the new angle is $\theta = \cos^{-1} 0.9 = 25.8°$. Therefore,

$$P_X = P_A \sin\theta = 133.3 \sin 25.8° = 58 \text{ kVAR}$$

The necessary correction is $58 - 120 = -62$ kVAR.

The new power triangle is shown in Fig. 7.8b. (Note that *apparent powers* cannot be combined algebraically.) The nearest commercial unit is 60 kVAR rated at 4000 V, and would be so specified. Since

$$P_X = V^2/X_C = V^2(-\omega C)$$

the capacitance for 60-Hz operation would be

$$\begin{aligned} C &= -P_X/\omega V^2 \\ &= 62 \times 10^3/377 \times 16 \times 10^6 \\ &= 10.3\,\mu\text{F}. \end{aligned}$$

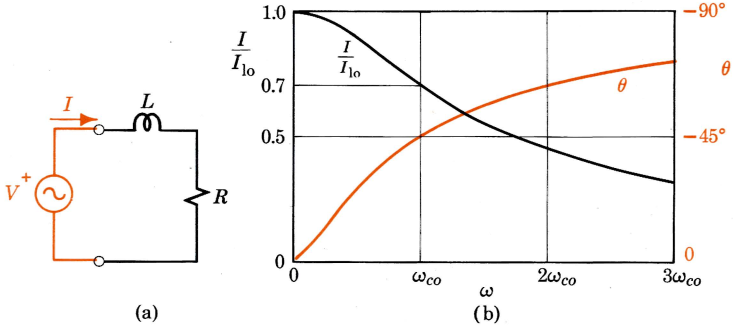
One plan for doing this is to include a power factor penalty clause in the rate schedule. The customer with low power factor pays a higher rate per kilowatt-hour (kWh). The usual cause of low power factor is lagging pf equipment such as induction motors, furnaces, and welders. If a power factor penalty is imposed, it may be economically justifiable to "correct" the situation by adding capacitors that draw a leading pf reactive power. Such capacitors operate at nearly zero power factor and are rated in kVAR at a specified voltage rather than in terms of the capacitance provided. (See Example 3.)

Frequency Response

The different responses of inductance and capacitance that permit power factor correction can be exploited in another way. Since $X_L = \omega L$ is directly proportional to frequency whereas $X_C = -1/\omega C$ is of opposite sign and inversely proportional to frequency, inductive and capacitive elements can be combined in networks that are *frequency selective.* For example, in a radio receiver a parallel combination of L and C is used to select a single signal from the myriad radio waves that impinge on the antenna. The extreme degree of frequency selectivity achievable is one of the key properties of electric circuits.

Simple Filters

The simple RL circuit of Fig. 7.9a is often employed to discriminate between input signals of different frequencies. For an input voltage of constant amplitude V and varying frequency $\omega = 2\pi f$, the impedance is $\mathbf{Z} = R + j\omega L$. At low frequencies, the current is limited by the resistance alone and $I_{lo} \cong V/R$. At high frequencies, the ωL term predominates and $I_{hi} \cong V/\omega L$ or current is inversely



(a) (b)

Figure 7.9 A frequency selective circuit.

proportional to frequency. In general,

$$\mathbf{I}(\omega) = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{\mathbf{V}}{R + j\omega L} = \frac{\mathbf{V}/R}{1 + j\omega L/R} \tag{7-12}$$

or

$$\frac{\mathbf{I}}{\mathbf{I}_{\text{lo}}} = \frac{1}{1 + j\omega L/R} = \frac{1}{\sqrt{1 + (\omega L/R)^2}} \underline{/\tan^{-1}(-\omega L/R)} = \frac{I}{I_{\text{lo}}} \underline{/\theta} \tag{7-13}$$

as shown in Fig. 7.9b. At the intermediate frequency

WL = R

$$\omega_{\text{co}} = \frac{R}{L} \tag{7-14}$$

resistive and reactive components of impedance are equal and

$$\frac{\mathbf{I}}{\mathbf{I}_{\text{lo}}} = \frac{1}{1 + j1} = \frac{1}{\sqrt{2}} \underline{/-45^\circ} = 0.707\underline{/-45^\circ} \tag{7-15}$$

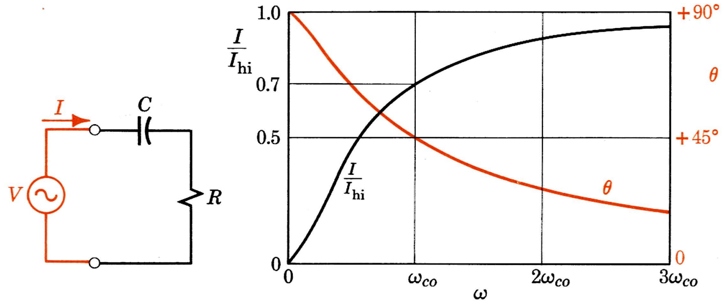
At this "70% current point", the power delivered to the circuit (I^2R) drops to one-half the maximum value. This *half-power frequency* is an easily calculated measure of the frequency response. Since signals of frequencies below ω_{co} are passed on to the resistance, this *RL* circuit is said to be a "low-pass filter" with a "cutoff frequency" of ω_{co} (see Exercise 14).

The simple *RC* circuit of Fig. 7.10a can serve as a high-pass filter, since at high frequencies the current is limited by the resistance alone and $I_{\text{hi}} = V/R$. In general,

$$\mathbf{I} = \frac{\mathbf{V}}{\mathbf{Z}} = \frac{\mathbf{V}}{R - (j1/\omega C)} = \frac{\mathbf{V}/R}{1 - (j1/\omega CR)} \tag{7-16}$$

or, in dimensionless form,

$$\frac{\mathbf{I}}{\mathbf{I}_{\text{hi}}} = \frac{1}{1 - (j1/\omega CR)} = \frac{1}{\sqrt{1 + (1/\omega CR)^2}} \underline{/\tan^{-1}(1/\omega CR)} = \frac{I}{I_{\text{hi}}} \underline{/\theta} \tag{7-17}$$



(a) (b)

Figure 7.10 A simple high-pass filter.

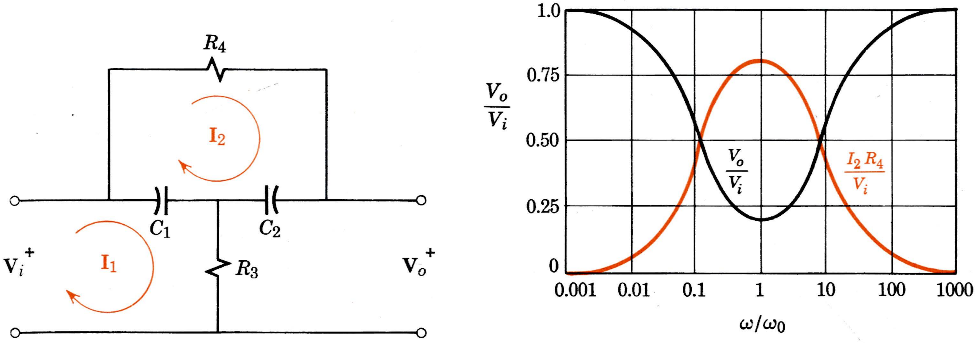
as shown in Fig. 7.10b. Here the cutoff or half-power frequency is defined by $(1/\omega CR) = 1$ or

$$\omega_{\text{co}} = \frac{1}{RC} \tag{7-18}$$

(Note that having analyzed the frequency response of series RL and RC circuits, we have also analyzed their duals. See Exercise 15.)

Bridged-T Filters

A more complicated frequency response is provided by the *bridged-T filter* of Fig. 7.11. At low frequencies, the capacitances are effectively open-circuits and the input voltage $\mathbf{V}_i$ is passed through via R_4. At high frequencies, C_1 and C_2 are effectively shorted and $\mathbf{V}_i$ is passed through directly. At an intermediate frequency ω_0 determined by the parameters of the circuit, the signal following one path is partially cancelled by the signal following the other and the output voltage $\mathbf{V}_o$ is reduced.



(a) Circuit

(b) Frequency response

Figure 7.11 The bridged-T filter.

In analyzing the response, it is convenient to find loop current $\mathbf{I}_2$ and express $\mathbf{V}_o$ as $\mathbf{V}_i - \mathbf{I}_2 R_4$. Assuming a high-impedance load so that the output current is negligible, we write

$$\Sigma\mathbf{V} = 0 = \mathbf{V}_i - \mathbf{I}_1(\mathbf{Z}_1 + \mathbf{Z}_3) + \mathbf{I}_2\mathbf{Z}_1 \tag{7-19}$$

$$\Sigma\mathbf{V} = 0 = \mathbf{I}_1\mathbf{Z}_1 - \mathbf{I}_2(\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_4) \tag{7-20}$$

From Eq. 7-20, $\mathbf{I}_1 = \mathbf{I}_2(\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_4)/\mathbf{Z}_1$; substituting in Eq. 7-19 yields,

$$\mathbf{I}_2 = \frac{\mathbf{V}_i}{\dfrac{\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_4}{\mathbf{Z}_1}(\mathbf{Z}_1 + \mathbf{Z}_3) - \mathbf{Z}_1} = \frac{\mathbf{V}_i}{\mathbf{Z}_2 + \mathbf{Z}_4 + \mathbf{Z}_3 + \dfrac{\mathbf{Z}_2\mathbf{Z}_3}{\mathbf{Z}_1} + \dfrac{\mathbf{Z}_3\mathbf{Z}_4}{\mathbf{Z}_1}} \tag{7-21}$$

Therefore,

$$\mathbf{I}_2\mathbf{Z}_4 = \frac{\mathbf{V}_i}{1 + \dfrac{\mathbf{Z}_3}{\mathbf{Z}_4} + \dfrac{\mathbf{Z}_2\mathbf{Z}_3}{\mathbf{Z}_1\mathbf{Z}_4} + \dfrac{\mathbf{Z}_2}{\mathbf{Z}_4} + \dfrac{\mathbf{Z}_3}{\mathbf{Z}_1}} = \frac{\mathbf{V}_i}{1 + \dfrac{\mathbf{Z}_3}{\mathbf{Z}_4}\left(1 + \dfrac{\mathbf{Z}_2}{\mathbf{Z}_1}\right) + \dfrac{\mathbf{Z}_2}{\mathbf{Z}_4}\left(1 + \dfrac{\mathbf{Z}_3\mathbf{Z}_4}{\mathbf{Z}_1\mathbf{Z}_2}\right)}$$

Expressing the impedances in terms of circuit parameters leads to

$$\mathbf{I}_2 R_4 = \frac{\mathbf{V}_i}{1 + \dfrac{R_3}{R_4}\left(1 + \dfrac{\omega C_1}{\omega C_2}\right) - j\dfrac{1}{\omega C_2 R_4}(1 - \omega^2 C_1 C_2 R_3 R_4)} \tag{7-22}$$

This expression becomes purely real for $\omega^2 C_1 C_2 R_3 R_4 = 1$ defining the critical frequency

$$\omega_0 = \frac{1}{\sqrt{C_1 C_2 R_3 R_4}} \tag{7-23}$$

The filter response is described by

$$\frac{\mathbf{V}_o}{\mathbf{V}_i} = 1 - \frac{\mathbf{I}_2 R_4}{\mathbf{V}_i} = 1 - \frac{1}{1 + \dfrac{R_3}{R_4}\left[1 + \dfrac{C_1}{C_2}\right] + j\dfrac{1}{\omega C_2 R_4}\left[\left(\dfrac{\omega}{\omega_0}\right)^2 - 1\right]} \tag{7-24}$$

By proper choice of the circuit parameters, a specified band of frequencies can be filtered out. (See Exercise 17.) The advantages of the bridged-T filter are its simplicity and low cost. If we are willing to use more complicated circuits and more expensive elements (inductances), we can obtain more complete filtering and sharper cutoff.†

Resonance

The slender suspension bridge across the Tacoma Narrows in Washington showed tendencies to oscillate up and down during construction and was nicknamed "Galloping Gertie." On November 7, 1940, only a few months after construction was completed, the oscillation began to build up under a moderate wind and then abruptly changed to a writhing motion. Within an hour the violent twisting had torn the 6-million dollar bridge to pieces. Ten years later, after much study, experiment, and wind-tunnel research, a new bridge designed to be stable in winds up to 120 mph was constructed on the same site.

This is an example of *resonance,* a phenomenon characteristic of low-loss second-order structures and systems. At a particular *resonant frequency* the impedance is small and the forced response may be very great even for moderate applied forces or voltages. There are many engineering applications of resonance. In an internal combustion engine with a "tuned" intake manifold, the resonance effect is used to increase the amount of fuel–air mixture delivered to the cylinders at high speed. In a refrigerator, the reciprocating compressor is mounted on a support designed to minimize the vibrations transmitted to the cabinet. In an electrical circuit with inductance and capacitance in series, there is always a frequency at which the two reactances just cancel resulting in the low-impedance characteristic of series resonance.

---

† Practical filters with sharper cutoff characteristics can be designed using more sophisticated techniques. See Chapter 19 of Skilling, *Electrical Engineering Circuits,* 2nd ed., John Wiley & Sons, New York, 1965.

Series Resonance

For a series RLC circuit the impedance is

$$\mathbf{Z} = R + j\omega L + \frac{1}{j\omega C} = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{7-25}$$

$$\mathbf{Z} = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \underline{/\tan^{-1}\frac{\omega L - 1/\omega C}{R}} \tag{7-26}$$

The frequency at which the reactances just cancel and the impedance is a pure resistance is the *resonant frequency* ω_0.

Example 4

The variable capacitor behind a radio dial provides a capacitance of 1 nF when the dial setting is "80." Calculate the inductance of the associated coil.

A dial setting of "80" means a frequency of 800 kilohertz (kHz). The coil-capacitor combination provides resonance at this frequency; therefore, $\omega_0 L = 1/\omega_0 C$ and

$$L = \frac{1}{\omega_0{}^2 C} = \frac{1}{(2\pi f_0)^2 C}$$

$$= \frac{1}{(2\pi \times 8 \times 10^5)^2 \times 1 \times 10^{-9}} = 39.6\ \mu\text{H}$$

In Fig. 7.12, where $\theta_Z = 0$ and $(\omega_0 L - 1/\omega_0 C) = 0$, resonant frequency is

$$\omega_0 = \frac{1}{\sqrt{LC}} \frac{\text{rad}}{\text{sec}} \quad \text{or} \quad f_0 = \frac{1}{2\pi\sqrt{LC}} \text{Hz} \tag{7-27}$$

By definition, at series resonance the impedance $Z_0 = R$ and $\theta_Z = 0$. At frequencies lower than ω_0, the capacitive reactance term predominates; the impedance increases rapidly as frequency decreases, and θ_Z approaches $-90°$. At frequencies higher than ω_0, the inductive reactance term predominates; the impedance increases rapidly with frequency and θ_Z approaches $+90°$. (Can you visualize the curve of θ_Z as a function of ω?)

Also shown in Fig. 7.12 is the variation with frequency of admittance $Y = 1/Z$. The admittance is high at resonance, where $Y_0 = 1/R$, and decreases rapidly as frequency is changed from ω_0. Since $I = YV$, the variation of current with

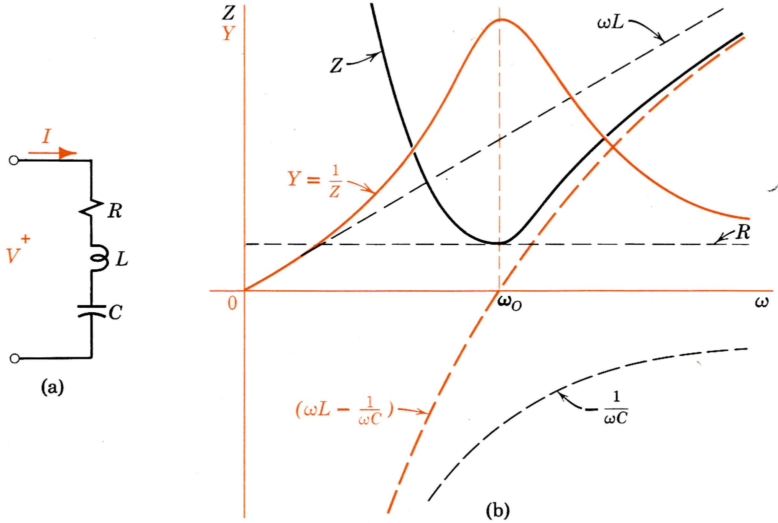


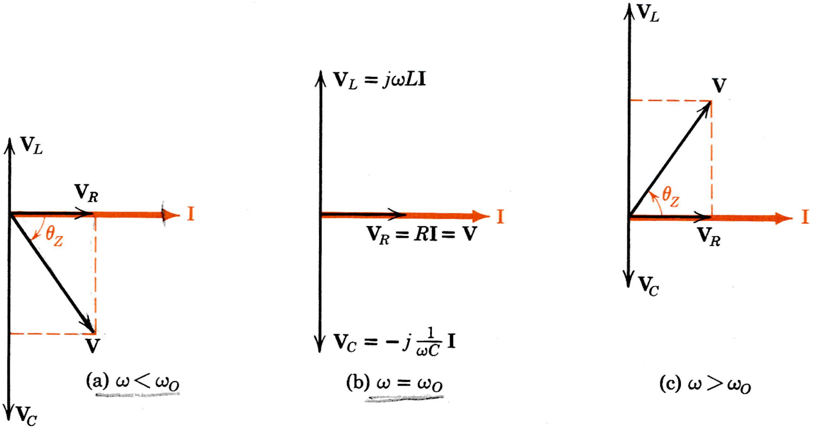
Figure 7.12 The frequency response of a series *RLC* circuit.

a constant amplitude voltage has the same shape as the admittance curve. If voltages of various frequencies are applied to a series *RLC* circuit, a voltage of frequency ω_0 will be favored and a relatively large current will flow in response to this particular voltage. The circuit is *frequency selective.*

Phasor Diagram Interpretation

Additional information about the phenomenon of resonance is revealed by phasor diagrams drawn for various frequencies. Since current is common to each element in this series circuit, phasor $\mathbf{I}$ (rms value) is drawn as a horizontal reference in Fig. 7.13. The voltage $\mathbf{V}_R$ across the resistance is in phase with the current, and voltages $\mathbf{V}_L$ and $\mathbf{V}_C$ across the inductance and capacitance, respectively, lead and lag the current. For $\omega = \omega_0$, $\mathbf{V}_L$ and $\mathbf{V}_C$ just cancel (Why?) and the voltage across the resistance is exactly equal to the applied voltage $\mathbf{V}$. For $\omega < \omega_0$, $\mathbf{V}_C$ is greater than $\mathbf{V}_L$ and the sum of the three voltages is $\mathbf{V}$. In comparison to the case for $\omega = \omega_0$, a larger $\mathbf{V}$ is required for the same current or the admittance is lower; θ_Z is negative indicating a leading power factor. For $\omega > \omega_0$, $\mathbf{V}_L$ is greater than $\mathbf{V}_C$ and the current lags $\mathbf{V}$ (the sum of the three component voltages) by angle θ_Z.

For a given current $\mathbf{I}$, the impedance $\mathbf{Z}$ is directly proportional to the voltage $\mathbf{V}$. Therefore, the voltage diagrams of Fig. 7.13 can be transformed into impedance diagrams. For $\omega < \omega_0$, the impedance is predominantly capacitive. For $\omega = \omega_0$, the impedance is a pure resistance and has a minimum value. For $\omega > \omega_0$, the impedance is predominantly inductive and increases with frequency.


$\mathbf{V}_L$
$\mathbf{V}_R$
I
θ_Z
V
$\mathbf{V}_C$
$\mathbf{V}_L = j\omega L\mathbf{I}$
$\mathbf{V}_R = R\mathbf{I} = \mathbf{V}$
$\mathbf{V}_C = -j\frac{1}{\omega C}\mathbf{I}$

(a) $\omega < \omega_0$ (b) $\omega = \omega_0$ (c) $\omega > \omega_0$

Figure 7.13 Phasor diagrams for a series *RLC* circuit.

Example 5

A generator supplies a variable frequency voltage of constant amplitude = 100 V (rms) to the circuit of Fig. 7.14 where $R = 5\ \Omega$, $L = 4$ mH, and $C = 0.1\ \mu$F. The frequency is to be varied until a maximum rms current flows. Predict the maximum current, the frequency at which it occurs, and the resulting voltages across the inductance and capacitance.

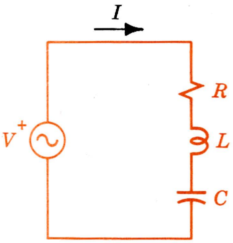


Figure 7.14 A series resonant circuit.

For a series *RLC* circuit, maximum current corresponds to maximum admittance which occurs at the resonant frequency. At $\omega = \omega_0$, $\omega_0 L = 1/\omega_0 C$, and $Y_0 = 1/Z_0 = 1/R$. Therefore,

$$I = VY = \frac{V}{R} = \frac{100}{5} = 20\ \text{A}$$

and

$$\omega_0 = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{4 \times 10^{-3} \times 10^{-7}}} = 5 \times 10^4\ \text{rad/sec}$$

Then

$$V_L = \omega LI = 5 \times 10^4 \times 4 \times 10^{-3} \times 20 = 4000\ \text{V}$$

and

$$V_C = \frac{I}{\omega C} = \frac{20}{5 \times 10^4 \times 10^{-7}} = 4000\ \text{V}$$

As expected, V_L is just equal to V_C and the phasor diagram is similar to that in Fig. 7.13b.

Quality Factor Q

How can the voltage across one series element be greater than the voltage across all three, as in Example 5? The answer is related to the fact that *L* and

C are energy-storage elements and high *instantaneous* voltages are possible. As indicated in Fig. 7.14b, phasors $\mathbf{V}_L$ and $\mathbf{V}_C$ are 180° out of phase; instantaneous voltages are also 180° out of phase, and a high positive voltage across L is canceled by a high negative voltage across C (see Exercise 20). The voltages do exist, however, and can be measured with ordinary voltmeters. The ability to develop high element voltages is a useful property of resonant circuits and we need a convenient measure of this property.

At resonance $\omega_0 L = 1/\omega_0 C$ and $I = I_0 = V/R$. Therefore the reactance voltages are

$$V_L = \omega_0 L I_0 = \omega_0 L \frac{V}{R} = \frac{\omega_0 L}{R} V \tag{7-28}$$

and

$$V_C = \frac{I_0}{\omega_0 C} = \frac{V}{\omega_0 C R} = \frac{\omega_0 L}{R} V = V_L \tag{7-29}$$

By definition, in a series circuit

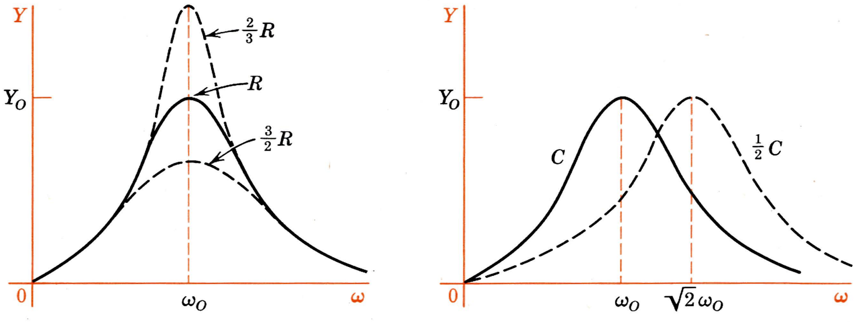
$$\frac{\omega_0 L}{R} = \frac{1}{\omega_0 C R} = Q \tag{7-30}$$

where Q is the *quality factor,* a dimensionless ratio.

In a practical circuit, R is essentially the resistance of the coil since practical capacitors have very low losses in comparison to practical inductors. Hence, Q is a measure of the energy-storage property (L) in relation to the energy dissipation property (R) of a coil or a circuit. A well-designed coil may have a Q of several hundred, and an RLC circuit employing such a coil will have essentially the same Q since the capacitor contributes very little to R. In practical coils, the losses (and therefore R in the linear model) increase with frequency and $Q = \omega L/R$ is fairly constant over a limited range of frequencies.

Equation 7-28 indicates that for a given value of current there is a "resonant rise in voltage" across the reactive elements equal to Q times the applied voltage. In addition to indicating the magnitude of this resonant rise, Q is a measure of the frequency selectivity of the circuit. A circuit with high Q discriminates sharply; a low-Q circuit is relatively unselective. The solid admittance curve of Fig. 7.15a is redrawn from Fig. 7.12; for the same L and C, the admittance curve near resonance depends on R. Increasing the resistance by 50% reduces Y_0 to $\frac{2}{3}$ of its original value and the selectivity is reduced. Decreasing the resistance to $\frac{2}{3}$ of its original value (and thereby increasing Q by a factor of $\frac{3}{2}$) increases Y_0 and increases selectivity.

The effect of changing one reactive element while keeping the other and the resistance constant is indicated by Fig. 7.15b. Although ω_0 is shifted, the general shape of the response curve is still the same. A *general resonance equation* that describes series RLC circuits with various parameters and various resonant



(a) Varying R (b) Varying C

Figure 7.15 The effect of circuit parameters on frequency response.

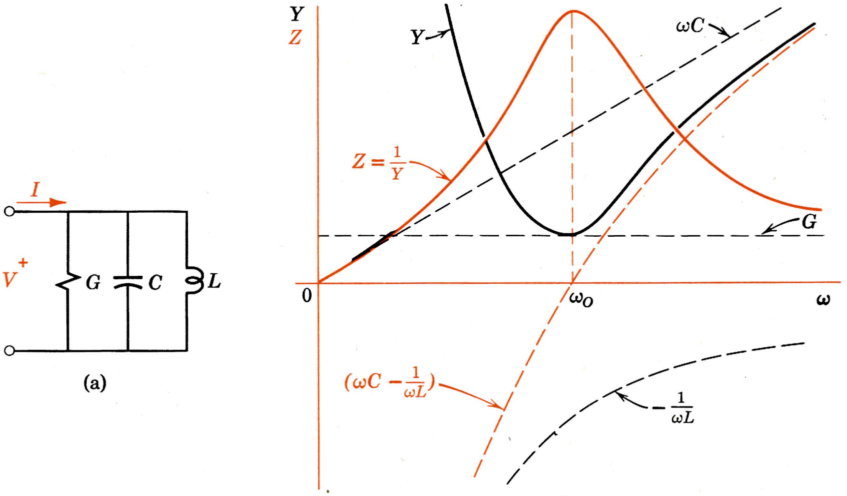
frequencies would be very valuable. Is it possible that such a general equation also describes parallel resonant circuits? Let us investigate that possibility.

Parallel Resonance

Fig. 7.16 shows a parallel GCL circuit. For this circuit the admittance is

$$\mathbf{Y} = G + j\omega C + \frac{1}{j\omega L} = G + j\left(\omega C - \frac{1}{\omega L}\right) \tag{7-31}$$

$$\mathbf{Y} = \sqrt{G^2 + \left(\omega C - \frac{1}{\omega L}\right)^2}\ \underline{/\tan^{-1}\frac{\omega C - 1/\omega L}{G}} \tag{7-32}$$



(a)

(b)

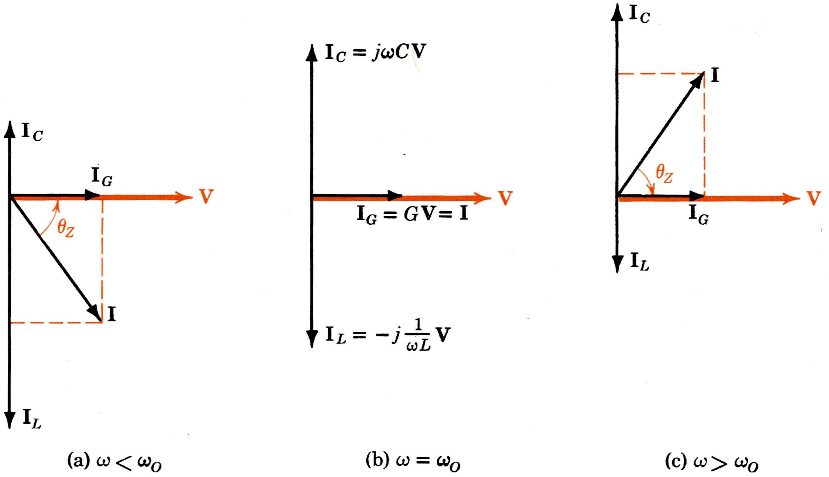
Figure 7.16 The frequency response of a parallel GCL circuit.

The frequency at which the susceptances just cancel and the admittance is a pure conductance is the resonant frequency ω_0. Where $\theta_Y = 0$ and $(\omega_0 C - 1/\omega_0 L) = 0$, the resonant frequency is

$$\omega_0 = \frac{1}{\sqrt{CL}} \frac{\text{rad}}{\text{sec}} \quad \text{or} \quad f_0 = \frac{1}{2\pi\sqrt{CL}} \text{Hz} \tag{7-33}$$

By definition, at parallel resonance the admittance $Y_0 = G$ and $\theta_Y = 0$. At frequencies lower than ω_0, the inductive susceptance term predominates; the admittance increases rapidly as frequency decreases and θ_Y approaches $-90°$. At frequencies higher than ω_0, the capacitive susceptance term predominates; the admittance increases rapidly with frequency and θ_Y approaches $+90°$.

These equations and descriptive statements look familiar; as a matter of fact, they were written by applying the duality transforms (Tables 5.2 and 5.3) to the corresponding equations (7-26 and 7-27) and statements for series resonance.† Since the circuits are duals, we expect duality in the frequency response curves; Fig. 7.16b is a relabeled version of Fig. 7.12b. (You may wish to translate for yourself the paragraph starting: "Also shown in Fig. 7.12 . . ." on p. 181.)


$\mathbf{I}_C$
$\mathbf{I}_G$
V
θ_Z
I
$\mathbf{I}_L$
$\mathbf{I}_C = j\omega C\mathbf{V}$
$\mathbf{I}_G = G\mathbf{V} = \mathbf{I}$
V
$\mathbf{I}_L = -j\frac{1}{\omega L}\mathbf{V}$
$\mathbf{I}_C$
I
θ_Z
V
$\mathbf{I}_G$
$\mathbf{I}_L$

(a) $\omega < \omega_0$ (b) $\omega = \omega_0$ (c) $\omega > \omega_0$

Figure 7.17 Phasor diagrams for a parallel *GCL* circuit.

The phasor diagrams for a parallel *GCL* circuit are shown in Fig. 7.17. In this case there is a "resonant rise in current" in the reactive elements. At resonance $V = V_0 = I/G$. Therefore the reactance currents are

---

† See Skilling, *Electrical Engineering Circuits*, 2nd ed., John Wiley & Sons, New York, 1965, pp. 232 ff.

$$I_C = \omega_0 C V_0 = \omega_0 C \frac{I}{G} = \frac{\omega_0 C}{G} I = Q_P I \tag{7-34}$$

and

$$I_L = \frac{V_0}{\omega_0 L} = \frac{I}{\omega_0 L G} = \frac{\omega_0 C}{G} I = Q_P I \tag{7-35}$$

Basic Definition of Q

The expressions for Q in a parallel circuit are quite different from those in a series circuit; in fact, if the same circuit elements are reconnected, $Q_P = 1/Q_S$. But basically the phenomena described by Q are the same in both circuits; apparently we need a more basic definition of Q. In essence, quality factor is a measure of the energy storage property of a circuit in relation to its energy dissipation property. A definition based on this concept is

$$Q = 2\pi \frac{\text{maximum energy stored}}{\text{energy dissipated per cycle}} \tag{7-36}$$

In a pendulum driven at its resonant frequency, the only energy input goes to supply friction losses; the energy stored is transferred back and forth from potential to kinetic. In the analogous electric circuit, the input impedance at resonance is a pure resistance and the stored energy is transferred back and forth between the magnetic field of the inductance and the electric field of the capacitance. It can be shown (see Problem 3) that at the resonant frequency the total stored energy is a constant. At an instant when the voltage across the capacitance is zero, the current in the inductance is maximum and all the stored energy is in the inductance. At an instant when v_C is maximum, i_L is zero and all the stored energy is in C. (Is this confirmed by the phasor diagrams of Figs. 7.13 and 7.17?)

Knowing the energy distribution, we can derive specific values of Q from the basic definition. For a series RLC circuit, the common current is $i = \sqrt{2}I \cos \omega t$, and all the stored energy is in L when $i = \sqrt{2}I$ or

$$W_{\text{stored}} = W_L = \tfrac{1}{2}Li^2 = LI^2$$

The energy dissipated per cycle is the energy per second (average power) divided by the frequency in hertz, or

$$W_{\text{diss/cycle}} = \frac{P}{f_0} = \frac{I^2 R}{f_0}$$

By Eq. 7-36,

$$Q_S = 2\pi \frac{LI^2}{I^2 R/f_0} = 2\pi f_0 \frac{L}{R} = \frac{\omega_0 L}{R} \tag{7-37}$$

Following a similar line of reasoning, the quality factor for a simple parallel circuit can be obtained (see Exercise 26). The great advantage of this basic definition of Q is that it is also applicable to more complicated lumped circuits, to distributed circuits such as transmission lines, and to nonelectrical systems. Note that for a reactive *component* such as a coil, $Q = \omega L/R$, a function of frequency; for a resonant *system*, Q is evaluated at ω_0 and is a constant.

Example 6

A typical radio receiver employs parallel resonant circuits in selecting the desired station. Using the values from Example 4 ($C = 1$ nF and $L = 39.6\ \mu$H), the resonant frequency is again 800 kHz and the dial setting is "80." If the GCL linear model has a Q of 100, what is the value of G?

If two signals of the same amplitude (50 μA rms) but different frequencies (f_1 at "80" and f_2 at "85") are introduced from the antenna, what are the respective voltages developed across the tuned circuit?

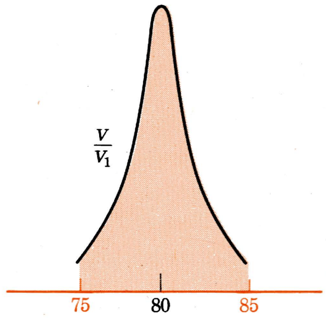


Figure 7.18 Frequency selectivity.

For a Q of 100, Eq. 7-34 yields

$$G = \frac{\omega_0 C}{Q_P} = \frac{2\pi \times 8 \times 10^5 \times 10^{-9}}{100} = 50.3\ \mu\mho$$

and the equivalent resistance is

$$R = \frac{1}{G} \cong 20{,}000\ \Omega$$

At the resonant frequency f_1,

$$V_1 = \frac{I_1}{Y_0} = \frac{I_1}{G} = \frac{50 \times 10^{-6}}{50.3 \times 10^{-6}} \cong 1\text{ V}$$

At frequency $f_2 = 850$ kHz, $\omega = 2\pi \times 8.5 \times 10^5$ rad/s,

$$\begin{aligned} Y_2 &= \sqrt{G^2 + (\omega C - 1/\omega L)^2} \\ &= \sqrt{(50.2)^2 + (5341 - 4728)^2} = 615\ \mu\mho \end{aligned}$$

and

$$V_2 = \frac{I_2}{Y_2} = \frac{50 \times 10^{-6}}{615 \times 10^{-6}} \cong 0.08\text{ V}$$

For the same impressed current, the voltage response for the desired signal is over 12 times as great as that for the undesired signal (Fig. 7.18), and the power developed in G is over 150 times as great since $P = V^2G$.

Normalized Response

The frequency selectivity demonstrated in Example 6 is characteristic of all resonant circuits and the response curves of all such circuits have the same general shape. Now we wish to derive a simple but general equation that describes series and parallel circuits, resonant at high and low frequencies, with large and small energy dissipation. To eliminate the effect of specific parameter values, we use dimensionless ratios. For illustration we choose to work with the admittance of a series *RLC* circuit, but the result has a more general interpretation. For the series circuit of Fig. 7.19a,

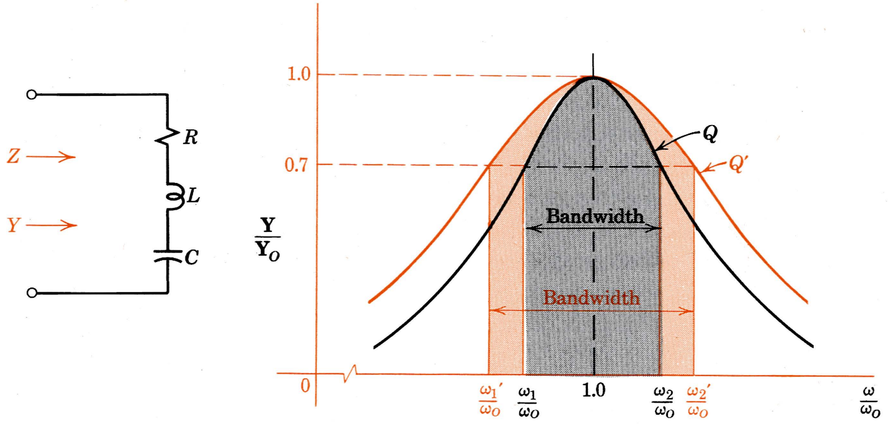
$$\mathbf{Z} = R + j\left(\omega L - \frac{1}{\omega C}\right) = \frac{1}{\mathbf{Y}}$$

and at resonance,

$$\mathbf{Z}_o = R = \frac{1}{\mathbf{Y}_o}$$

The dimensionless ratio of the admittance at any frequency ω to that at the resonant frequency ω_o is

$$\frac{\mathbf{Y}}{\mathbf{Y}_o} = \frac{\mathbf{Z}_o}{\mathbf{Z}} = \frac{R}{R + j\left(\omega L - \dfrac{1}{\omega C}\right)} = \frac{1}{1 + j\left(\dfrac{\omega L}{R} - \dfrac{1}{\omega CR}\right)} \tag{7-38}$$



(a) (b)

Figure 7.19 The normalized response of a series *RLC* circuit.

Introducing the factor ω_o/ω_o and letting the dimensionless ratio $\omega_o L/R = Q = 1/\omega_o CR$, the admittance ratio for a series *RLC* circuit is

$$\frac{\mathbf{Y}}{\mathbf{Y}_o} = \frac{1}{1 + j\left(\dfrac{\omega L}{R} \cdot \dfrac{\omega_o}{\omega_o} - \dfrac{1}{\omega CR} \cdot \dfrac{\omega_o}{\omega_o}\right)} = \frac{1}{1 + jQ\left(\dfrac{\omega}{\omega_o} - \dfrac{\omega_o}{\omega}\right)} \tag{7-39}$$

This simple equation also describes the impedance ratio $\mathbf{Z}/\mathbf{Z}_o$ for a parallel *GCL* circuit; the even simpler reciprocal equation describes the impedance ratio for a series circuit and the admittance ratio for a parallel circuit.

Bandwidth

The curves in Fig. 7.19b have been normalized; all variables are dimensionless ratios. The difference in the two curves shown is due to a difference in selectivity. A convenient quantitative measure of selectivity is defined by letting the imaginary term in the denominator of Eq. 7-39 be equal to $\pm j1$. Where

$$\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right) = \pm\frac{1}{Q}$$

then

$$\frac{\mathbf{Y}}{\mathbf{Y}_o} = \frac{1}{1 \pm j1} = \frac{1}{\sqrt{2}} \;\underline{/\mp 45^\circ} \qquad (7\text{-}40)$$

Imposing this condition defines two frequencies, ω_1 and ω_2, such that

$$\frac{\omega_1}{\omega_o} - \frac{\omega_o}{\omega_1} = -\frac{1}{Q} \quad \text{and} \quad \frac{\omega_2}{\omega_o} - \frac{\omega_o}{\omega_2} = +\frac{1}{Q} \qquad (7\text{-}41)$$

Frequencies ω_1 and ω_2 are called the lower and upper "70% points," since at these frequencies the magnitude Y/Y_o is $1/\sqrt{2} = 0.707$. For a given applied voltage, current is proportional to admittance and these are also called "70% current points." More generally, these are *half-power frequencies* because power is proportional to the square of the current. (Do ω_1 and ω_2 also designate half-power frequencies in parallel circuits?)

The frequency range between the half-power points, $\omega_2 - \omega_1$, is called the *bandwidth*, a convenient measure of the selectivity of the circuit. Solving Eqs. 7-41 and selecting the consistent roots, we obtain

$$\omega_1 = \omega_o\sqrt{1 + \left(\frac{1}{2Q}\right)^2} - \frac{\omega_o}{2Q} \quad \text{and} \quad \omega_2 = \omega_o\sqrt{1 + \left(\frac{1}{2Q}\right)^2} + \frac{\omega_o}{2Q} \qquad (7\text{-}42)$$

Example 7

A series circuit is to be resonant at 800 kHz. Specify the value of C required for a given inductor ($L = 40\ \mu\text{H}$ and $R = 4.02\ \Omega$), and predict the bandwidth.

For the specified resonant frequency,

$$C = \frac{1}{\omega_o^2 L} = \frac{1}{(2\pi 8 \times 10^5)^2 \times 4 \times 10^{-5}} = 0.99\ \text{nF}$$

and the Q of the circuit or the inductor is

$$Q = \frac{\omega_o L}{R} = \frac{2\pi 8 \times 10^5 \times 4 \times 10^{-5}}{4.02} = 50$$

Then the bandwidth is, by Eq. 7.43,

$$f_2 - f_1 = \frac{f_o}{Q} = \frac{800 \times 10^3}{50} = 16\ \text{kHz}$$

Subtracting the first equation from the second, the bandwidth is

$$\omega_2 - \omega_1 = \frac{\omega_0}{Q} \quad \text{or} \quad f_2 - f_1 = \frac{f_0}{Q} \tag{7-43}$$

Note that the bandwidth is inversely proportional to the quality factor Q, a very convenient result. (See Example 7.)

In the typical case where $Q \geq 10$, the factor $\sqrt{1 + (1/2Q)^2} \cong 1$ and

$$\omega_1 = \omega_0 - \frac{\omega_0}{2Q} \quad \text{and} \quad \omega_2 = \omega_0 + \frac{\omega_0}{2Q} \tag{7-44}$$

In other words, the resonance curve is symmetric and the half-power points are equidistant from the resonant frequency, or

$$\omega_2 - \omega_0 = \omega_0 - \omega_1 = \frac{\omega_0}{2Q} \tag{7-45}$$

This simple relation makes it easy to sketch the response curve of a series or parallel resonant circuit by plotting three points.

Practical Resonant Circuits

In focusing attention on principles, we neglected some practical aspects of resonant circuits. The practical parallel circuit consists of an inductor in parallel with a capacitor, and the only resistance is usually that due to losses in the inductor. The linear model of Fig. 7.20 is the appropriate representation of the actual circuit; if Q is 20 or more, the general equations (Eqs. 7-38 through 45) hold. Example 8 illustrates practical circuit design.

The emphasis in this discussion is on a variable frequency ω. In many cases the circuit is "tuned" by varying a capacitor or, less frequently, an inductor. This corresponds to varying ω_0 in Eq. 7-39 and the same general resonance behavior is observed.

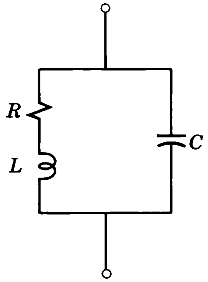
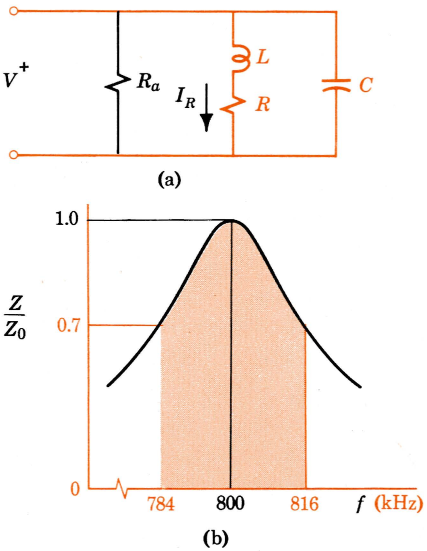


Figure 7.20 Model of a practical parallel circuit.

The specific values of the circuit parameters are usually determined by economic considerations. At audio frequencies (20 to 20,000 Hz), inductors with iron cores are used to obtain inductances in the order of henrys. At radio broadcast frequencies (above 500,000 Hz), air-core inductors with inductances in the order of microhenrys are resonated with variable air-dielectric capacitors with capacitances in the order of nanofarads. Reasonable values of resistance and quality factor are obtained in properly designed inductors using copper wire. At higher frequencies, undesired resonance may occur because of the inductance of a short lead (a small fraction of a microhenry) and the capacitance between two adjacent conductors (a few picofarads).

Example 8

Design a parallel circuit to be resonant at 800 kHz with a bandwidth of 32 kHz. Use the inductor from Example 7 with $L = 40\ \mu\text{H}$ and $R = 4.02\ \Omega$, and predict the impedance at resonance.



(a)

(b)

Figure 7.21 Design of a parallel resonant circuit.

The desired quality factor is

$$Q = \frac{f_0}{f_2 - f_1} = \frac{800 \times 10^3}{32 \times 10^3} = 25$$

Since $Q > 20$, Eq. 7-38 holds for this practical circuit and $C = 1/\omega_0^2 L = 0.99$ nF as before. From Example 7, the Q of the inductor is 50. Therefore, to double the bandwidth we must halve the Q and double the losses. One possibility is to add a 4.02-Ω resistance in series with R.

Another possibility is shown in Fig. 7-21. To double the losses, add R_a so that

$$P_a = \frac{V^2}{R_a} = I_R^2 R \cong \left(\frac{V}{\omega_0 L}\right)^2 R$$

or

$$R_a = \frac{(\omega_0 L)^2}{R} \cdot \frac{R}{R} = Q^2 R = 50^2 \times 4.02 = 10{,}050\ \Omega$$

The parallel combination of inductor, capacitance, and resistance can be replaced by the model of Fig. 7.16a where (Eqs. 7-32 and 7-35),

$$G = Y_0 = \frac{1}{Z_0} = \frac{1}{Q\omega_0 L}$$

Therefore,

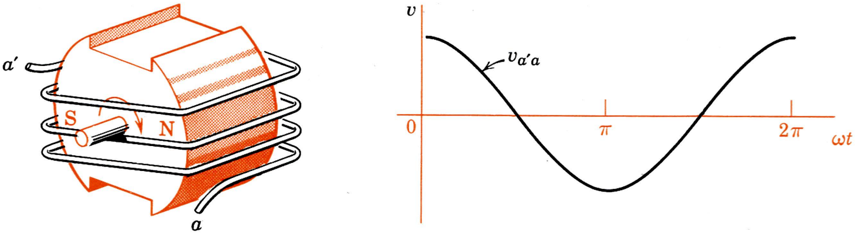
$$Z_0 = Q\omega_0 L = 25 \times 2\pi 8 \times 10^5 \times 4 \times 10^{-5} = 5025\ \Omega$$

Three-Phase Circuits†

The generation and transmission of electrical power is more efficient in polyphase systems employing combinations of two, three, or more sinusoidal voltages. In addition, polyphase circuits and machines possess some unique advantages; for example, power in a three-phase circuit is constant rather than pulsating as it is in a single-phase circuit. Also three-phase motors start and run much better than single-phase motors. The most common form of polyphase system employs three *balanced* voltages, equal in magnitude and differing in phase by 360°/3 = 120°. The discussion here is restricted to balanced three-phase circuits. First we see how such voltages are generated and connected, and then we learn to analyze the resulting circuits.

---

† This section can be postponed and considered in connection with polyphase machines in Chapter 25.

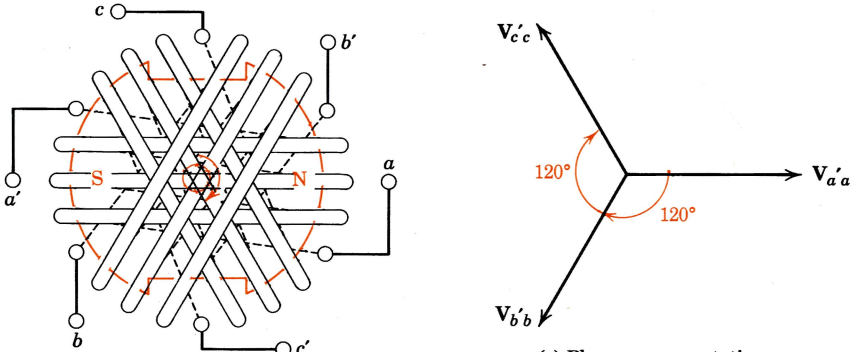


(a) Elementary generator (b) Voltage generated in the winding

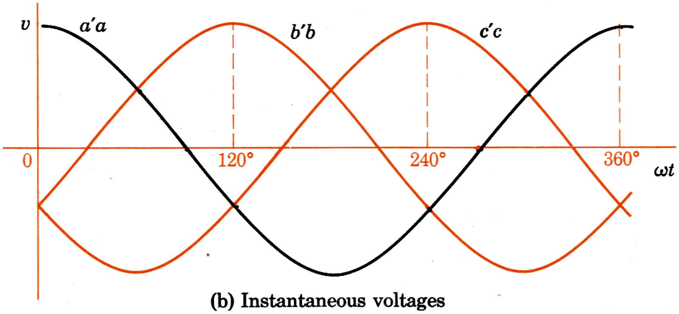
Figure 7.22 Generation of an alternating voltage.

Three-Phase Voltage Generation

The elementary ac generator of Fig. 7.22 consists of a rotating magnet and a stationary winding. The turns of the winding are spread along the periphery of the machine. The voltage generated in each turn of the four-turn winding is slightly out of phase with the voltage generated in its neighbor because it is cut by maximum magnetic flux density an instant earlier or later. (Machine construction and voltage generation are discussed in Chapter 23.) The voltages in the four turns are in series and, therefore, they add to produce voltage $v_{a'a}$.



(a) Winding arrangement (c) Phasor representation



(b) Instantaneous voltages

Figure 7.23 Balanced three-phase voltages.

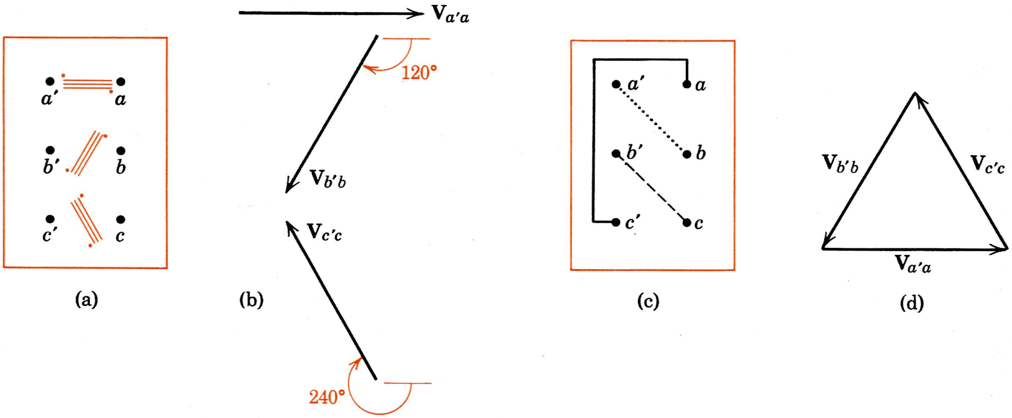
If the winding were continued around the machine, the voltage generated in the last turn would be 180° out of phase with that in the first and they would cancel, producing no useful effect. For this reason, a winding is commonly spread over no more than one-third of the periphery; the other two-thirds can be used to generate two other similar voltages. The three sinusoids (sinusoids are obtained with a proper winding distribution and magnet shape) generated by the three similar windings are shown in Fig. 7.23. Using double-subscript notation to minimize confusion, we describe the voltages as:

$$\begin{aligned} v_{a'a} &= \sqrt{2}V\cos\omega t & \mathbf{V}_{a'a} &= V\underline{/0^\circ} \\ v_{b'b} &= \sqrt{2}V\cos(\omega t - 120^\circ) & \mathbf{V}_{b'b} &= V\underline{/-120^\circ} \\ v_{c'c} &= \sqrt{2}V\cos(\omega t - 240^\circ) & \mathbf{V}_{c'c} &= V\underline{/-240^\circ} \end{aligned} \tag{7-46}$$

The three similar portions of a three-phase system are called "phases," a slightly different use of the word. Because the voltage in phase *a′a* reaches its maximum first, followed by that in phase *b′b*, and then by that in phase *c′c*, we say the *phase rotation* is *abc*. This is an arbitrary convention; for any given machine the phase rotation may be reversed by reversing the direction of rotation of the magnet or by interchanging the labels on two of the three-phase windings. We shall assume the phase rotation is *abc* unless otherwise stated.

Delta Connection

Three separate *single-phase* loads can be connected to the three windings of a generator and supplied with power independently. The three-phase system running along the rear property lines in residential areas could be operated in this way. The first customer could be connected across phase *a*, the second across phase *b*, and the third across phase *c*. However, savings in wire and other benefits are gained by interconnecting the three phases.



(a) (b) (c) (d)

Figure 7.24 Three-phase generator and Δ connection.

Consider the terminal board of a three-phase generator as shown in Fig. 7.24a and the corresponding voltage phasors. If terminal c' is connected to terminal a (solid line in Fig. 7.24c), voltage $a'c$, equal to voltage $a'a$ plus voltage $c'c$, appears across terminals $a'c$. In phasor notation,

$$\mathbf{V}_{a'c} = \mathbf{V}_{a'a} + \mathbf{V}_{c'c}$$

If then terminal b' is connected to terminal c (as shown by the dashed line), the voltage appearing across terminals $a'b$ is

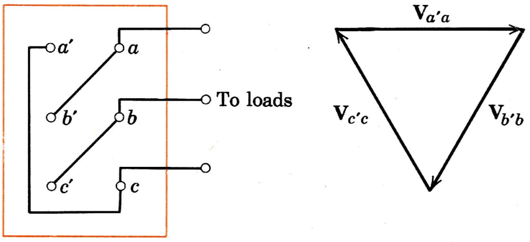
$$v_{a'b} = v_{a'a} + v_{c'c} + v_{b'b}$$

or

$$\mathbf{V}_{a'b} = \mathbf{V}_{a'a} + \mathbf{V}_{c'c} + \mathbf{V}_{b'b} \tag{7-47}$$

The graph of instantaneous values in Fig. 7.23b shows that $v_{a'b} = 0$ and this is verified by the sum of the balanced phasors in Fig. 7.23c. (The sum of a set of symmetric complex quantities is always zero. See Exercise 34.)

Since the voltage between terminals a' and b is zero, these terminals can be connected (dotted line) and the result is a system of voltages connected in *delta*. The same three windings can be reconnected as in Fig. 7.25 to yield another delta connection. In either Δ (delta), a voltage appears across each pair of terminals; three wires connected to terminals a, b, and c can supply three single-phase loads. This is the arrangement commonly employed to serve residential customers. Can three such wires carry as much power as six wires of the same size supplying separate single-phase loads? (See Problem 9.)



(a) (b)

Figure 7.25 Another Δ connection.

Wye Connection

The same machine can be reconnected to give a quite different result. If terminals a', b', and c' are connected together (Fig. 7.26a), a, b, and c become the output terminals of a *wye* or *star* connection. Connecting a, b, and c together would yield another **Y** (wye). In the Δ connection, the voltage across the output terminals is just equal to the voltage of a winding. In the **Y** connection, however,

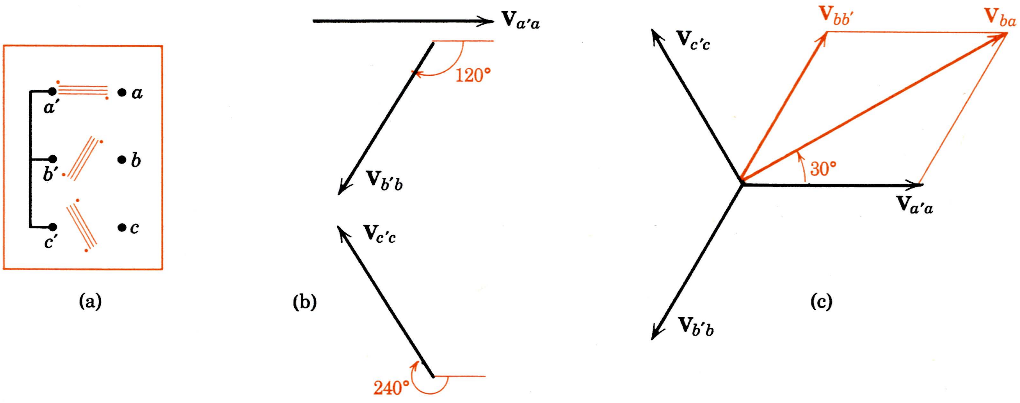

a
a′
b
b′
c
c′
(a)
$\mathbf{V}_{a'a}$
120°
$\mathbf{V}_{b'b}$
$\mathbf{V}_{c'c}$
240°
(b)
$\mathbf{V}_{c'c}$
$\mathbf{V}_{bb'}$
$\mathbf{V}_{ba}$
30°
$\mathbf{V}_{a'a}$
$\mathbf{V}_{b'b}$
(c)

Figure 7.26 Three-phase generator and Y connection.

the voltage across terminals *ba* is (from Fig. 7.26c)

$$\mathbf{V}_{ba} = \mathbf{V}_{a'a} - \mathbf{V}_{b'b} = \mathbf{V}_{a'a} + \mathbf{V}_{bb'} = 2\mathbf{V}_{a'a}\cos 30^\circ \underline{/30^\circ} = \sqrt{3}\mathbf{V}_{a'a}\underline{/30^\circ}$$

In general,

$$\mathbf{V}_{\text{line}} = \sqrt{3}\mathbf{V}_{\text{phase}}\underline{/30^\circ} \tag{7-48}$$

or, in words, *in a* Y *connection the line-to-line voltage is* $\sqrt{3}$ *times the phase voltage and is displaced* 30° *in phase.*

Delta-Circuit Calculations

Given a three-phase source and load, a common problem is to calculate the various currents and voltages and the power supplied to the load. In Fig. 7.27 the load consists of three equal impedances connected in **Δ**. The internal connection of the source is unimportant; it could be Δ or Y. The only requirement is that there be available at terminals *a*, *b*, and *c* three balanced voltages.

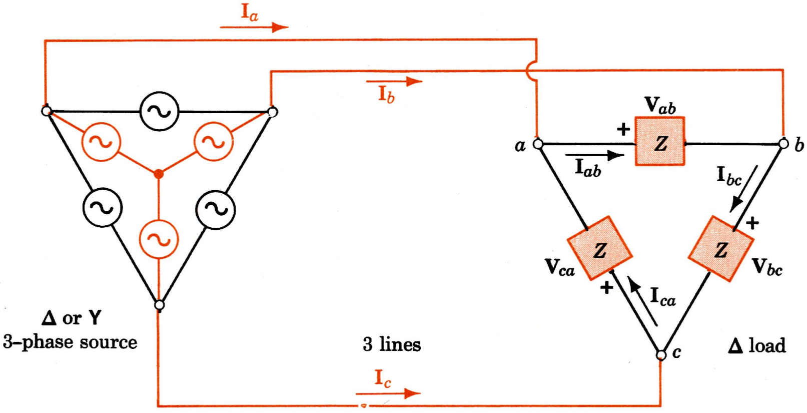
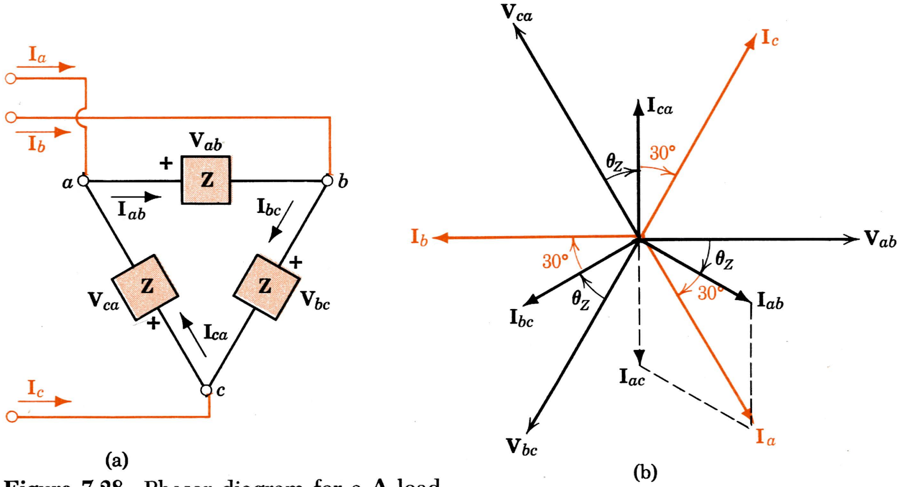

$\mathbf{I}_a$
$\mathbf{I}_b$
$\mathbf{V}_{ab}$
+
Z
a
b
$\mathbf{I}_{ab}$
$\mathbf{I}_{bc}$
+
Z
Z
$\mathbf{V}_{ca}$
$\mathbf{V}_{bc}$
+
$\mathbf{I}_{ca}$
Δ or Y
3-phase source
3 lines
c
Δ load
$\mathbf{I}_c$

Figure 7.27 A delta-connected load with **Δ** or **Y** source.

Although three-phase circuits are no more difficult to solve than single-phase circuits, there are many opportunities for confusion in dealing with three of everything. A clearly labeled schematic wiring diagram should be drawn and a phasor diagram should be used as a guide and a check. Calculations are simplified and mistakes reduced if, at first, the wiring diagram and the phasor diagram have the same orientation. In Fig. 7.28, if $\mathbf{V}_{ab}$ is taken as the horizontal reference, phasor $\mathbf{V}_{ab}$ is in the same direction as a line going from *a* to *b*.† Phasor $\mathbf{V}_{bc}$ is in the same direction as a line from *b* to *c* and is 120° "behind" $\mathbf{V}_{ab}$ (Fig. 7.28b). Similarly, $\mathbf{V}_{ca}$ is 240° behind $\mathbf{V}_{ab}$.



(a) (b)

Figure 7.28 Phasor diagram for a **Δ** load.

In a delta load the voltages across the three impedances (the phase voltages) are identical with the voltages between the "lines" from source to load (the line voltages). The currents in each phase (labeled consistently with the phase voltages in Fig. 7.28a) are obtained in the customary way. Using phasor representation, the *phase* currents are

$$
\begin{aligned}
\mathbf{I}_{ab} &= \frac{\mathbf{V}_{ab}}{\mathbf{Z}_{ab}} = \frac{V\angle 0^\circ}{Z\angle \theta_Z} = \frac{V}{Z}\angle 0^\circ - \theta_Z \\
\mathbf{I}_{bc} &= \frac{\mathbf{V}_{bc}}{\mathbf{Z}_{bc}} = \frac{V\angle -120^\circ}{Z\angle \theta_Z} = \frac{V}{Z}\angle -120^\circ - \theta_Z \\
\mathbf{I}_{ca} &= \frac{\mathbf{V}_{ca}}{\mathbf{Z}_{ca}} = \frac{V\angle -240^\circ}{Z\angle \theta_Z} = \frac{V}{Z}\angle -240^\circ - \theta_Z
\end{aligned}
\tag{7-49}
$$

---

† In general v_{ab} is the voltage of *a* with respect to *b*. If Z is a pure resistance, v_{ab} and i_{ab} are directly proportional or "in phase."

The *line* currents are not equal to the phase currents. For example,

$$\mathbf{I}_a = \mathbf{I}_{ab} + \mathbf{I}_{ac} = \mathbf{I}_{ab} - \mathbf{I}_{ca}$$

This addition is indicated in Fig. 7.28b. Since $\mathbf{I}_{ac} = -\mathbf{I}_{ca}$,

$$\mathbf{I}_a = 2I_{ab} \cos 30^\circ \underline{/-\theta_Z - 30^\circ} = \sqrt{3}\mathbf{I}_{ab}\underline{/-30^\circ}$$

$\mathbf{I}_b$ and $\mathbf{I}_c$ are obtained similarly and the result is a set of three balanced line currents. In general,

$$\mathbf{I}_{\text{line}} = \sqrt{3}\mathbf{I}_{\text{phase}}\underline{/-30^\circ} \qquad (7\text{-}50)$$

or, in words, *in a* **Δ** *connection the line current is* $\sqrt{3}$ *times the phase current and is displaced* -30° *in phase.*

Example 9

A load consisting of three identical impedances $\mathbf{Z} = 10\underline{/-45^\circ}\ \Omega$ in **Δ** is connected to a three-phase, 220-V source. Determine phase and line currents, and draw a labeled phasor diagram.

The source is assumed to supply balanced three-phase voltages with a line-to-line value of 220 V rms and phase rotation *abc*. If $\mathbf{V}_{ab}$ is taken as the horizontal reference, a suitable wiring diagram is shown in Fig. 7.29a. The phase currents are labeled consistently in a clockwise direction and the line currents are all assumed *into* the load to preserve the balance.

Obviously, the phase voltages are just equal to the line voltages. The phase currents are

$$\mathbf{I}_{ab} = \frac{\mathbf{V}_{ab}}{\mathbf{Z}} = \frac{220\underline{/0^\circ}}{10\underline{/-45^\circ}} = 22\underline{/+45^\circ}\text{ A}$$

$$\mathbf{I}_{bc} = \frac{\mathbf{V}_{bc}}{\mathbf{Z}} = \frac{220\underline{/-120^\circ}}{10\underline{/-45^\circ}} = 22\underline{/-75^\circ}\text{ A}$$

$$\mathbf{I}_{ca} = \frac{\mathbf{V}_{ca}}{\mathbf{Z}} = \frac{220\underline{/-240^\circ}}{10\underline{/-45^\circ}} = 22\underline{/-195^\circ}\text{ A}$$

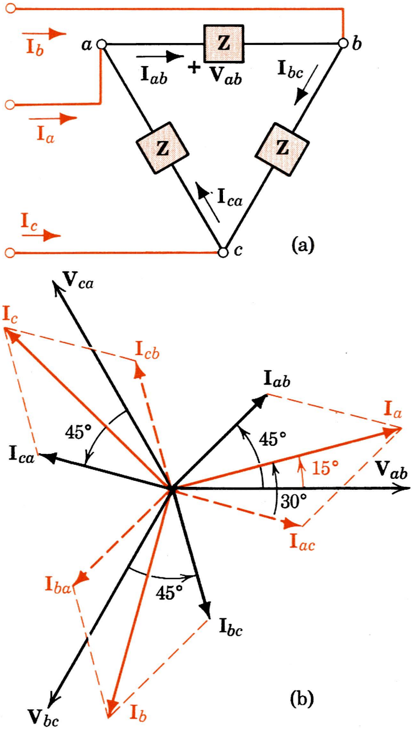
The current in line *a* is

$$\mathbf{I}_a = \mathbf{I}_{ab} + \mathbf{I}_{ac} = 22\underline{/+45^\circ} + 22\underline{/-15^\circ}$$
$$= 22\sqrt{3}\underline{/45^\circ - 30^\circ} = 38\underline{/+15^\circ}\text{ A}$$

By symmetry,

$$\mathbf{I}_b = 38\underline{/-105^\circ}\text{ A} \quad \text{and} \quad \mathbf{I}_c = 38\underline{/-225^\circ}\text{ A}$$

The voltages, phase currents, and line currents are shown on the phasor diagram in Fig. 7.29b.



$\mathbf{I}_b$
a
Z
b
$\mathbf{I}_{ab}$
+ $\mathbf{V}_{ab}$
$\mathbf{I}_{bc}$
$\mathbf{I}_a$
Z
Z
$\mathbf{I}_{ca}$
$\mathbf{I}_c$
c
(a)
$\mathbf{V}_{ca}$
$\mathbf{I}_c$
$\mathbf{I}_{cb}$
$\mathbf{I}_{ab}$
$\mathbf{I}_a$
45°
45°
$\mathbf{I}_{ca}$
15°
$\mathbf{V}_{ab}$
30°
$\mathbf{I}_{ac}$
$\mathbf{I}_{ba}$
45°
$\mathbf{I}_{bc}$
$\mathbf{V}_{bc}$
$\mathbf{I}_b$
(b)

Figure 7.29 Delta load calculations.

Power Calculations

The total power in a balanced three-phase load is the sum of three equal phase powers or

$$P_{\text{total}} = 3P_p = 3V_pI_p \cos\theta \tag{7-51}$$

where $\cos\theta$ is the power factor of the load or θ is the angle between phase voltage V_p and phase current I_p. As illustrated by Example 9, θ is *not* the angle between line voltage and line current. It is easier to measure line quantities, however, so an expression for total power in terms of V_l and I_l is useful.

In a Δ load, $V_l = V_p$ and $I_l = \sqrt{3}I_p$; therefore,

$$P_\Delta = 3V_pI_p \cos\theta = 3V_l\frac{I_l}{\sqrt{3}}\cos\theta = \sqrt{3}V_lI_l \cos\theta \tag{7-52}$$

In a Y load, $V_l = \sqrt{3}V_p$ and $I_l = I_p$; therefore,

$$P_Y = 3V_pI_p \cos\theta = 3\frac{V_l}{\sqrt{3}}I_l \cos\theta = \sqrt{3}V_lI_l \cos\theta \tag{7-53}$$

The same expression holds for power in a Δ or a Y connected load (see Example 10 on p. 200).

The measurement of power in a practical situation is not straightforward because phase currents in a Δ-connected load and phase voltages in a Y-connected load are usually not accessible. However, some clever engineer observed that two wattmeters, properly connected, can measure the total power in any three-phase system, balanced or unbalanced. Referring to Fig. 7.30a, it is clear that three wattmeters with their current coils connected in series with the lines and their voltage coils connected from line to neutral would measure the three phase powers and their sum would be the total power. It can be shown that, since we are interested only in potential differences, the location of the common terminal or "neutral" of the three voltage coils is unimportant; the sum of the three wattmeter readings is always the total power. If now the "neutral" of the wattmeter coils is connected to line *c*, one wattmeter will sense no voltage therefore it will read zero and can be dispensed with.

If the *two-wattmeter method* is applied in Example 10, one wattmeter W_1 reads the product of $\mathbf{I}_a$ and $\mathbf{V}_{ac}$ (*not* $\mathbf{V}_{ca}$) and the cosine of the angle between them or

$$W_1 = I_a\, V_{ac} \cos\theta_1 = 12.7 \times 220 \times \cos 15^\circ = 2698 \text{ W}$$

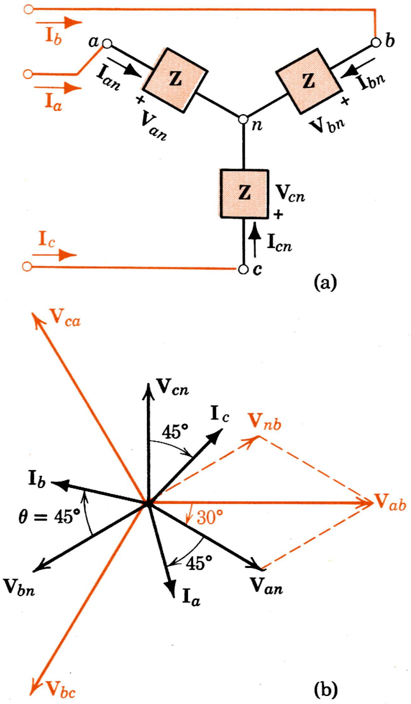
and the other wattmeter reads

$$W_2 = I_b\, V_{bc} \cos\theta_2 = 12.7 \times 220 \times \cos 75^\circ = 723 \text{ W}$$

and the total is $W_1 + W_2 \cong 3420$ W as calculated. (The derivation of the general relation is left for the curious student; see Problem 11.)

Example 10

Three equal impedances of $\mathbf{Z} = 10\underline{/+45^\circ}\ \Omega$ are connected in Y across a 220-V supply. Determine phase voltages, phase and line currents, and phase and total power. Draw a labeled phasor diagram.



$\mathbf{I}_b$
a
$\mathbf{I}_{an}$
Z
+
$\mathbf{V}_{an}$
$\mathbf{I}_a$
n
Z
$\mathbf{I}_{bn}$
b
+
$\mathbf{V}_{bn}$
Z
$\mathbf{V}_{cn}$
+
$\mathbf{I}_{cn}$
$\mathbf{I}_c$
c
(a)
$\mathbf{V}_{ca}$
$\mathbf{V}_{cn}$
45°
$\mathbf{I}_c$
$\mathbf{V}_{nb}$
$\mathbf{I}_b$
θ = 45°
30°
$\mathbf{V}_{ab}$
45°
$\mathbf{V}_{bn}$
$\mathbf{I}_a$
$\mathbf{V}_{an}$
$\mathbf{V}_{bc}$
(b)

Figure 7.30 Power calculations.

A properly oriented wiring diagram is shown in Fig. 7.30a. By inspection, line current I_a = phase current I_{an}. (Point n is called the "neutral" and is important in unbalanced systems.) From Eq. 7-48,

$$\mathbf{V}_p = \frac{\mathbf{V}_l}{\sqrt{3}}\underline{/-30^\circ} \quad \text{or} \quad \mathbf{V}_{an} = \frac{\mathbf{V}_{ab}}{\sqrt{3}}\underline{/-30^\circ}$$

In this problem,

$$\mathbf{V}_{an} = (220/\sqrt{3})\underline{/0^\circ - 30^\circ} = 127\underline{/-30^\circ}\text{ V}$$

This is consistent with the orientation of the wiring diagram; the other phase voltages are determined by inspection and drawn on the phasor diagram as

$$\mathbf{V}_{bn} = 127\underline{/-150^\circ}\text{ V} \quad \text{and} \quad \mathbf{V}_{cn} = 127\underline{/-270^\circ}\text{ V}$$

Then

$$\mathbf{I}_a = \mathbf{I}_{an} = \frac{\mathbf{V}_{an}}{\mathbf{Z}_{an}} = \frac{127\underline{/-30^\circ}}{10\underline{/+45^\circ}} = 12.7\underline{/-75^\circ}\text{ A}$$

By symmetry,

$$\mathbf{I}_b = \mathbf{I}_{bn} = 12.7\underline{/-195^\circ}\text{ A}$$

and

$$\mathbf{I}_c = \mathbf{I}_{cn} = 12.7\underline{/-315^\circ}\text{ A}$$

as shown on the phasor diagram. The phase power is

$$P_p = V_p I_p \cos\theta = 127 \times 12.7(0.707) = 1140\text{ W}$$

The total power is, by Eq. 7-53,

$$\begin{aligned} P_t &= \sqrt{3}V_l I_l \cos\theta \\ &= \sqrt{3} \times 220 \times 12.7(0.707) \cong 3420\text{ W} \end{aligned}$$

Check:

$$P_t = 3P_p = 3 \times 1140 \cong 3420\text{ W}$$

As illustrated in the Examples, polyphase circuits are complicated, and clearly labeled wiring and phasor diagrams are necessary to minimize confusion. In a given phase, calculations are handled just as in single-phase circuits. The symmetry of balanced three-phase systems reduces the work required.

SUMMARY

- For sinusoids: Power (average) $P = VI\cos\theta = I_R{}^2R = V_R{}^2/R$ in W

  Reactive power $P_X = VI\sin\theta = I_X{}^2X = V_X{}^2/X$ in VAR

  Apparent power $P_A = VI = I^2Z = V^2/Z$ in VA

  Complex power $\mathbf{P}_A = P + jP_X$

  Power factor $= \cos\theta = P/VI$

  Reactive factor $= \sin\theta = P_X/VI$

- To determine total apparent power, add power and reactive power components separately and combine vectorially; display on power triangles.

- A circuit containing inductance and capacitance is in resonance if the terminal voltage and current are in phase. At the resonant frequency the power factor is unity, and the impedance and admittance are purely real.

  Resonant frequency: $\omega_0 = 1/\sqrt{LC}$ or $f_0 = 1/2\pi\sqrt{LC}$

- The frequency selectivity of a resonant circuit is determined by bandwidth.

$$BW = \omega_2 - \omega_1 = \frac{\omega_0}{Q} \quad \text{where} \quad Q = 2\pi\frac{\text{maximum energy stored}}{\text{energy dissipated per cycle}}$$

For $Q \geq 10$, $\omega_1 = \omega_0 - \dfrac{\omega_0}{2Q}$ and $\omega_2 = \omega_0 + \dfrac{\omega_0}{2Q}$

- For resonant circuits:

$$\text{Series:} \quad Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR} \quad \text{and} \quad V_L = V_C = QV \text{ at } \omega_0$$

$$\text{Parallel:} \quad Q = \frac{\omega_0 C}{G} = \frac{1}{\omega_0 LG} \quad \text{and} \quad I_L = I_C = QI \text{ at } \omega_0$$

$$\text{Series } \frac{\mathbf{Y}}{\mathbf{Y}_0} = \text{parallel } \frac{\mathbf{Z}}{\mathbf{Z}_0} = \frac{1}{1 + jQ\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)}$$

- A balanced three-phase system consists of three equal single-phase sources connected in **Δ** or **Y** supplying three equal loads connected in **Δ** or Y. For balanced three-phase systems in

  Δ: $\mathbf{V}_{\text{line}} = \mathbf{V}_{\text{phase}}$ and $\mathbf{I}_{\text{line}} = \sqrt{3}\mathbf{I}_{\text{phase}}\underline{/-30^\circ}$

  Y: $\mathbf{I}_{\text{line}} = \mathbf{I}_{\text{phase}}$ and $\mathbf{V}_{\text{line}} = \sqrt{3}\mathbf{V}_{\text{phase}}\underline{/+30^\circ}$

- In analyzing balanced three-phase circuits:

  Draw a carefully oriented and clearly labeled wiring diagram.

  Sketch the phasor diagram as a guide and as a check.

  Analyze one phase and use symmetry for the other phases.

  Total power $= 3V_{\text{phase}}I_{\text{phase}}\cos\theta = \sqrt{3}V_{\text{line}}I_{\text{line}}\cos\theta$.

REVIEW QUESTIONS

1. What is the frequency of the power pulsations in a 60-Hz circuit?
2. What is the interpretation of "negative power p" delivered "to" a passive circuit?
3. What is the interpretation of "negative reactive power P_X"?
4. Sketch a sinusoidal voltage, a sinusoidal current leading by about 45°, and the corresponding instantaneous power curve.
5. What is measured in VA? In VAR?
6. Is power a vector or a scalar quantity? What is "complex power"?
7. Outline the steps in adding two apparent powers.
8. Why is the local public utility interested in the pf of a plant?
9. Cite an example of resonance from acoustics, aeronautics, and hydraulics.
10. In terms of energy storage, what is the condition necessary for resonance?
11. What actually happens when you turn the tuning dial of a radio?
12. How is the resonant frequency ω_0 related to the undamped natural frequency ω_n?
13. Above ω_0, is a series RLC circuit capacitive or inductive? Below ω_0?
14. How can the voltage across one series element in an RLC circuit be greater than the total voltage across all three?
15. Why is Q called the "quality factor"?
16. Using the duality transformations, translate the next full paragraph after Eq. 7-27.
17. How could you determine the Q of a pendulum experimentally?
18. Name two important virtues of resonant circuits.
19. Do ω_1 and ω_2 defined by Eq. 7-41 designate half-power frequencies in parallel circuits?
20. Why is Y/Y_0 called "relative response"?
21. If $\mathbf{I}_1 = I\underline{/30^\circ}$ is one current in a balanced three-phase system, what are the other two?
22. Given $\mathbf{V}_{12} = V\underline{/90^\circ}$, define the other two voltages of a balanced three-phase system and show them connected in Δ and Y.
23. Explain by phasor diagrams the $\sqrt{3}$ in the V and I relations in Y and Δ circuits.
24. In analyzing a Δ load, why is it unimportant whether the source is Δ or Y?
25. In a "440-V, three-phase" system, what voltage is 440 V? Is this rms?
26. What is the advantage of using the same orientation for wiring and phasor diagrams?

EXERCISES

1. When a current $i = 2.82 \cos 500t$ A flows in a device, the voltage developed is $v = 141 \cos (500t - 60^\circ)$ V.
   (a) Sketch current and voltage waveforms approximately to scale.
   (b) Sketch the instantaneous power on the graph of part (a).
   (c) Compute the average power and the reactive power and show on graph.
   (d) Determine the impedance of the device and specify the series components of a linear model.
2. In the circuit of Fig. 7.31, ammeter AM reads 10 A and voltmeter VM reads 250 V.
   (a) Where in this circuit is average power dissipated? Predict the reading on wattmeter WM.
   (b) Predict total reactive power and overall power factor. Leading or lagging?
3. In the circuit of Fig. 7.32, $R_1 = 25\,\Omega$,

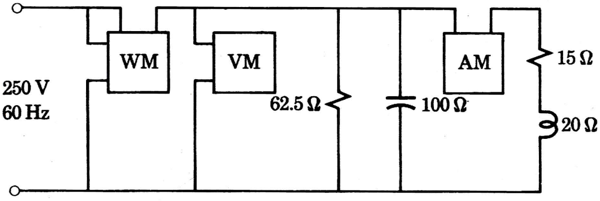


Figure 7.31

$R_2 = 24\ \Omega$, and $X_L = 18\ \Omega$. Wattmeter WM reads 4000 W and ammeter AM reads 8 A.

(a) Find I_2 and V_2.

(b) Using V_2 as a horizontal reference, show V_2, I_2, and I_C on a phasor diagram.

(c) Show I_1 on the phasor diagram and determine X_C.

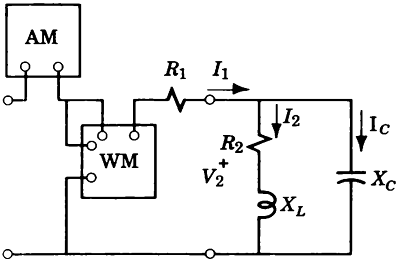


Figure 7.32

4. In the circuit of Fig. 7.32, $R_1 = 10\ \Omega$, $R_2 = 20\ \Omega$, and $X_L = 20\ \Omega$. If $I_2 = 10$ A and $I_C = 10$ A,
   (a) Predict the readings on the ammeter and the wattmeter.
   (b) Calculate the reactive powers in X_L and X_C.
5. Predict the reading of a wattmeter in which:
   (a) a current $i = 10\sqrt{2}\cos(\omega t + 45°)$ A flows in the current coil and a voltage $v = 220\sqrt{2}\cos(\omega t - 15°)$ V is applied to the voltage coil.
   (b) a current $i = 5\cos(\omega t - 120°)$ A flows in the current coil and a voltage $v = 20\cos(\omega t - 75°)$ V is applied to the voltage coil.
6. A load connected across a 200-V, 60-Hz line draws 10 kW at a lagging pf of 0.5.
   (a) Determine the current and the reactive power.
   (b) What series combination of circuit elements is equivalent to this load?
   (c) What parallel combination of circuit elements is equivalent to this load?
7. Repeat Exercise 6 for a load drawing 20 kW at 0.5 pf leading.
8. Meters are connected to read the current and power into a load impedance **Z** from a 200-V, 60-Hz source.
   (a) Draw the wiring diagram.
   (b) If the ammeter reads 5 A and the wattmeter reads 600 W, determine the power, reactive power, apparent power, and power factor.
   (c) When a small pure capacitive reactance X_C is connected in series with **Z**, the ammeter reads 4 A. Is **Z** capacitive or inductive? Predict the new wattmeter reading.
9. An industrial load consists of 100 kVA of induction motors operating at 0.6 pf lagging and 20 kW of resistance heating. Sketch the total plant power diagram. Calculate the total kVA and the plant power factor.
10. Specify the reactive kVA of capacitors required to correct the power factor of the plant in Exercise 9 to 0.85 lagging. If this plant is supplied by a transformer at voltage V, compare the magnitudes of the I^2R losses in the transformer in the original and corrected cases.
11. On a 2000-V (rms) line, there is an inductive load, drawing 40 kW and 60 kVAR, in parallel with a capacitive load drawing 40 kVA at 0.6 pf.
    (a) Express the total complex power in rectangular and polar form.
    (b) Show each load and the total on a power diagram.
    (c) Calculate the total line current.
    (d) Specify the unit to be connected in parallel with the two loads to "correct" the power factor to 0.95.
12. A typical industrial rate schedule includes a penalty for low power factor. List two reasons why the utility is concerned about your power factor.
13. An industrial load consists of the following: 10 kW of lighting at 100 V, a resistance heating unit that draws 150 A at 200 V, and an induction furnace that requires 360 kW and 400 kVAR at 2000 V. Power is supplied to the plant at 40 kV and 60 Hz.
    (a) Draw a complex power triangle for the total plant load and predict the input current and power factor.
    (b) Specify the capacitor (connected in parallel with the load) required to correct the power factor to 0.8.
14. For a load resistance $R = 10\ \text{k}\Omega$, design a simple *RL* low-pass filter with a cutoff frequency at $f = 1000$ Hz.

15. (a) Draw the dual of Fig. 7.9a and sketch the frequency-response curves; describe the frequency-response characteristics of the new circuit in words.
(b) Repeat for the dual of Fig. 7.10a.

16. The circuit of Fig. 7.33 is one type of filter. The input consists of four sinusoidal components with $\omega_1 = 10$, $\omega_2 = 100$, $\omega_3 = 1000$, and $\omega_4 = 10{,}000$ rad/s.
(a) Derive a general expression for the transfer ratio V_o/V_i.
(b) For $C = 10\,\mu F$, $L = 1$ H, and $R = 10\,k\Omega$, evaluate the ratio V_o/V_i for each component and draw a conclusion.

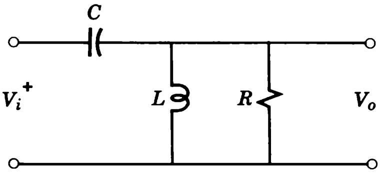


Figure 7.33

17. In the bridged-T filter of Fig. 7.11, $C_2 = 5\,\mu F = 5\,C_1$ and $R_4 = 5\,k\Omega = 5\,R_3$. Analyze the frequency response, i.e., determine ω_0 and calculate just enough points to sketch V_o/V_i vs ω/ω_0.

18. Examine the dial of a radio and determine the frequency range of the AM broadcast band. Calculate the range of capacitance needed with a 100-μH inductance to tune over the band.

19. Write the duals of: "impedance," "reactance," "resistance," "inductance," and "capacitance." Write out a statement (50 words or less) describing the behavior of a series resonant circuit including each of the above words, underlined. Obtain a statement of the behavior of a parallel resonant circuit by substituting the duals.

20. A 10-nF capacitor is connected in series with a coil of 2 mH inductance and 5 Ω resistance.
(a) Determine the resonant frequency.
(b) If a voltage of 2 V is to appear across the capacitor at resonance, what total voltage must be applied to the series combination?
(c) How can the voltage across one series component be larger than the total applied voltage? Sketch v_C and v_L as functions of ωt.

21. A voltage source with a constant magnitude of 10 V (rms) and a variable frequency is applied to a series combination of capacitance C and inductance $L = 25$ mH with a Q of 20 over a fairly wide range of frequency. The maximum value of current is observed at $\omega = 80$ krad/s. Calculate:
(a) the resonant frequency in hertz,
(b) the maximum value of current (rms),
(c) the values of R and C,
(d) the voltage across C at resonance.

22. A resonant circuit consists of a capacitance C in series with an inductor of inductance L and resistance R. Voltage V is applied.
(a) Define the resonant frequency ω_0 and the quality factor Q.
(b) Derive an expression for V_C at ω_0 in terms of Q and V.
(c) Derive an expression for the voltage across the inductor at ω_0 in terms of Q and V.
(d) Draw a conclusion regarding the relative magnitudes of V_C and V_{LR} for $Q > 10$.
(e) Derive an expression for the phase angle between inductor voltage and current in terms of Q.

23. Design a series RLC circuit to have a bandwidth of $10/\pi$ kHz and a resonant frequency of $500/\pi$ kHz. Use an available inductor that has an inductance of 0.4 mH and a Q of 80.

24. Design a parallel resonant circuit to provide an impedance at resonance of 50,000 Ω, a resonant frequency of 2 MHz, and a Q of 80.

25. In Fig. 7.34, $I = 2$ mA at $\omega = 5 \times 10^6$ rad/s, $G = 5\,\mu\mho$, and $L = 2$ mH.
(a) For what value of C is the voltage V a maximum?
(b) Predict the magnitude of this maximum voltage.
(c) For $I = 2$ mA at $\omega = 2 \times 10^6$ rad/s (C unchanged), predict V.
(d) Under the condition of part (a), predict the magnitude of I required to establish a current of 1 A in L.

26. (a) Using the basic definition, derive an

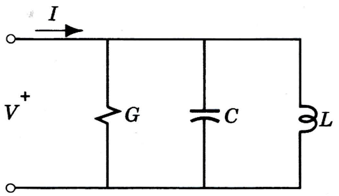


Figure 7.34

expression for Q for the circuit of Fig. 7.34.

(b) Show that, in any resonant circuit, at the half-power frequencies the resistance is just equal to the magnitude of the reactance.

27. The power supplied by a signal generator to an RLC series circuit is measured by a wattmeter. The signal generator output is 2 V at all frequencies. As the frequency is varied, a maximum average power of 4 W is observed at a frequency of $10/2\pi$ kHz. At frequencies $9.5/2\pi$ and $10.5/2\pi$ kHz, the power is observed to be 2 W.

(a) Draw the circuit including signal generator and wattmeter.

(b) Determine resonant frequency ω_0, quality factor Q, and bandwidth.

(c) Determine R, L, and C.

28. A series circuit has a resonant frequency of 1000 Hz; at 1050 Hz the power factor is 0.707. Calculate the Q.

29. A coil has a reactance of 2000 Ω and a resistance of 40 Ω at 2500 kHz. It is tuned to series resonance at 2500 kHz with a capacitor. Plot a few easily determined points and sketch a curve of current for 1 V applied as a function of frequency over a range from approximately 2400 to 2600 kHz.

30. In the practical tuning circuit of Fig. 7.20, $R = 20\ \Omega$, $L = 5$ mH, and $C = 0.5\ \mu$F.

(a) Determine the frequency at which the equivalent series impedance is a pure resistance.

(b) For the frequency of part (a), sketch a phasor diagram showing $\mathbf{V}_C$, $\mathbf{V}_R$, $\mathbf{V}_L$, $\mathbf{I}_C$, $\mathbf{I}_R$, and total $\mathbf{I}$.

(c) For the frequency of part (a), determine the equivalent series resistance.

31. For the parallel circuit of Fig. 7.20,

(a) Derive an expression for admittance as a function of ω.

(b) For $R = 25\ \Omega$ and $L = 0.05$ mH, specify C for resonance at 10^7 rad/s.

(c) For your circuit, estimate the Q, estimate ω_0 following the suggestion on p. 191, compare to the specified ω_0, and draw a conclusion.

32. Design a parallel circuit to have a bandwidth of $25/\pi$ kHz and a resonant frequency of $1000/\pi$ kHz. The available inductor has an inductance of 0.4 mH and a Q of 80. Draw a labeled circuit diagram. Estimate the impedance of your circuit at the resonant frequency. State assumptions made.

33. The terminals of a three-phase generator (Fig. 7.25a) are numbered from 1 to 6. Measurements indicate the voltages are as follows:

$$v_{12} = 100\sqrt{2}\cos\omega t \text{ V}$$
$$v_{34} = 100\sqrt{2}\cos(\omega t + 60°) \text{ V}$$
$$v_{56} = 100\sqrt{2}\cos(\omega t + 120°) \text{ V}$$

(a) Draw a delta voltage system using these voltages, clearly indicating what connections are to be made on the panel and which are the output terminals.

(b) Repeat for a wye voltage system.

34. Given the circuit of Fig. 7.35 with:

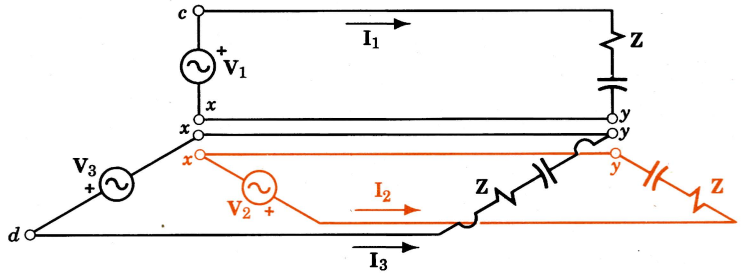


Figure 7.35

$\mathbf{V}_1 = 240\underline{/90°}$ V $\quad$ $\mathbf{V}_3 = 240\underline{/-150°}$ V

$\mathbf{V}_2 = 240\underline{/-30°}$ V $\quad$ $\mathbf{Z} = 40\underline{/-30°}$ Ω

(a) Calculate phasors $\mathbf{I}_1$, $\mathbf{I}_2$, and $\mathbf{I}_3$.
(b) If terminals x are connected together and terminals y are connected together, what will be current $\mathbf{I}_{yx}$?
(c) Under the condition of part (b), what will be the voltage $\mathbf{V}_{cd}$?
(d) Draw the phasors of $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}_3$, on a single phasor diagram with all phasors starting from a common point.

35. (a) Three wires enter a building. The voltages with respect to ground are measured to be: 130, 0, and 130 V. The voltage between the two outside wires is 260 V. Describe this electrical distribution system in terms your roommate would understand.
(b) Repeat part (a) if the three voltages with respect to ground are 130, 150, and 130 V, and the voltage between any pair of wires is 260 V.

36. Three equal impedances are connected in **Δ** across a 240-V line. One of the three line-to-line voltages is $240\underline{/60°}$ V and one of the phase currents is $6\underline{/200°}$ A.
(a) Identify the other voltages and phase currents and show all on a phasor diagram (all phasors starting from the same point).
(b) Determine the phase impedance in complex form.
(c) Determine the three line currents in phasor form.

37. Three equal impedances of $25\underline{/+60°}$ Ω are connected in **Δ** across a 400-V line.
(a) Draw a clearly labeled wiring diagram properly oriented.
(b) Calculate phase and line currents.
(c) Draw a labeled phasor diagram of voltages and currents.
(d) Calculate the total power.

38. Repeat Exercise 37 with the impedances connected in Y.

39. Repeat Exercise 37 for three equal impedances $\mathbf{Z} = 8 - j6$ Ω connected in **Δ.**

40. In urban areas it is desirable to supply "110-V" appliances and "220-V" appliances from the same bank of transformers. A convenient compromise is a Y-connected source with the neutral grounded. If a line-to-line voltage of 208 V is acceptably close to "220," what voltage is available from line to ground?

41. Repeat Exercise 39 with the impedances connected in Y.

42. Design a three-phase, 30-kW heater for operation on a 220-V line by specifying the resistance and power rating of the three identical resistors connected in Y.

43. Repeat Exercise 42 assuming **Δ**-connected resistors.

44. A balanced **Δ** capacitive load is connected as shown in Fig. 7.28a. A voltmeter across line *ab* reads 300 V. An ammeter in line *a* reads 20 A. A wattmeter with voltage coil across *ab* and current coil in line *a* reads 4240 W. Sketch the phasor diagram and determine the complex phase impedance **Z.**

45. A 4000-V, three-phase system supplies two balanced three-phase loads: a 300-kVA (input) Y-connected induction motor operating at 0.6 pf and a 120-kW **Δ**-connected heater. Predict the total line current. (Hint: Does the type of internal connection affect the line current?)

46. For the load in Example 9:
(a) Redraw Fig. 7.29a to show the WM connections in a power measurement by the two-wattmeter method.
(b) Draw the phasors of the voltages and currents sensed by the meters.
(c) Predict the WM readings and compare to the result of Eq. 7-52.

PROBLEMS

1. The monthly rate schedule for a given plant is: first 5000 kWh @ 5¢/kWh, next 20,000 kWh @ 4¢, next 75,000 kWh @ 3¢, and excess at 2¢. The total charge for energy (kWh) is decreased or increased by 0.5% for each 1% that the average pf is greater or less than 85%.

For a certain manufacturing plant operating 16 hours per day the monthly consumption is 180,000 kWh and 160,000 kVARh (lagging).

(a) Compute the average monthly "power" bill.

(b) Analyze a proposal by the electrical engineer to improve plant pf to 85% by adding capacitors ($10 per kVAR with 25% annual fixed charges covering depreciation, maintenance, etc.).

2. An industrial load is represented in Fig. 7.36 by $R = 6\ \Omega$ and $X_L = 8\ \Omega$. The load voltage **V** is $250\underline{/0^\circ}$ V.

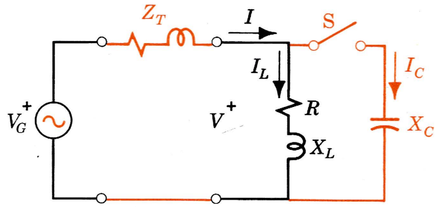


Figure 7.36

(a) Calculate the load current, power, reactive power, and power factor.

(b) Calculate the generator voltage $\mathbf{V}_G$ required at the input end of the transmission line (represented by the series impedance $\mathbf{Z}_T = 1 + j3\ \Omega$) and the power lost in transmission P_T.

(c) If capacitor $X_C = 12.5\ \Omega$ is connected in parallel by closing switch S, calculate $\mathbf{I}_C$, the new load current **I**, and the new power factor. Show **V**, $\mathbf{I}_L$, $\mathbf{I}_C$, and **I** on a phasor diagram.

(d) Calculate the new generator voltage and the new transmission power loss.

(e) What two advantages do you see for "improving the power factor" by adding a parallel capacitor?

3. Prove that the energy stored in a series *RLC* circuit excited at its resonant frequency is a constant and not a function of time.

4. In terms of displacement, the natural response of a pendulum is

$$x = X_m e^{-0.1t} \cos 5t$$

Define and determine the Q of this pendulum.

5. A circuit consists of capacitance C in series with a parallel combination of resistance R and inductance L. Determine the impedance $Z(j\omega)$ between the terminals at resonance, derive an expression for the resonant frequency ω_0, and determine the Q at resonance.

6. Energy is stored in an RLC circuit in the form of an initial voltage V_0 on the capacitor. Using the basic definition of Q, derive expressions for the *damping coefficient* α and the *natural frequency* ω in terms of Q and the resonant frequency ω_0. Rewrite Eq. 4-21 in terms of Q and ω_0. (Hint: Approximate the stored energy after one cycle and use the first two terms of $e^\theta = 1 + \theta + \cdots$)

7. In a parallel GCL circuit, $L = 2$ mH and $Q = 50$ at $f_0 = 400$ kHz.

(a) Define resonance for this circuit and determine the value of C.

(b) Determine the half-power frequencies and the bandwidth of the resonant circuit.

(c) If a total current $I = 50\ \mu$A flows into the parallel circuit, predict the voltage across the circuit.

(d) The practical parallel resonant circuit consists of a capacitor C in parallel with an inductor with inductance $L = 2$ mH and resistance R. Estimate the value of R at $f = 400$ kHz if the Q of the practical circuit is 50.

8. For the practical tuning circuit of Fig. 7.20.

(a) Define the resonant frequency ω_0.

(b) Derive a general expression for ω_0 in terms of $Q = \omega L/R$ and ω_S, the resonant frequency for the same elements in series.

9. For the same line-to-line voltage and the same current per conductor, compare the power transmission capacity of six wires supplying three single-phase resistive loads with the capacity of three wires supplying one three-phase resistive load. Draw a conclusion regarding the economics of three-phase systems.

10. Three identical impedances $Z_Y\underline{/\theta_Y}$ are connected in Y across a balanced three-phase line. Determine the magnitude Z_Δ and phase angle θ_Δ of the impedances which when connected in Δ across the same line would draw the same power and reactive power.

11. Demonstrate the validity of the two-wattmeter method of power measurement. First redraw Fig. 7.30a and show the two meters properly connected. Then draw phasors $\mathbf{V}_{ac}$, $\mathbf{V}_{bc}$, $\mathbf{V}_{an}$, $\mathbf{V}_{bn}$, $\mathbf{I}_a$, and $\mathbf{I}_b$ for any power factor angle θ. Finally, express the wattmeter readings in general terms and show that their sum agrees with Eqs. 7-52 and 53.

12. When power is measured in a balanced three-phase load, the two wattmeter readings are 3.8 kW and 1.9 kW. What is the power factor of the load?

Chapter 8

General Network Analysis

One-Port Networks
Two-Port Networks

In analyzing specific circuits, we developed some techniques that can be applied to circuits in general. Important principles that are broadly applicable to circuits of considerable complexity we called *network theorems*. In Chapter 2 we used some of these principles to simplify one-port resistive networks and to draw general conclusions about their behavior. In this chapter we extend those theorems to the case of ac circuits and then introduce two new theorems and show how they are applied. As before, we demonstrate the validity of the concepts and leave general proofs to more advanced courses.

Another important class of circuits includes *two-port* or *three-terminal* networks. Transformers and amplifiers are two-ports; energy or information is introduced at the *input* and abstracted at the *output*. The last half of this chapter is devoted to the analysis and application of some common two-port networks.

One-Port Networks

The theorems applied in Chapter 2 to two-terminal resistive networks are easily extended to include general ac one-ports.

Equivalence

Two one-ports are equivalent if they present the same v-i characteristics.

Two passive one-ports are equivalent if they have the same input impedance or the same input admittance; for sinusoidal excitation and response, for example, this means the same **Z** or **Y**. Ordinarily, two real networks are equivalent at one frequency only.

Network Reduction

Replacing a complicated network with a relatively simple equivalent is advantageous in network analysis. Already we have used the fact that in passive circuits impedances and admittances are readily combined. For impedances in series,

$$\mathbf{Z}_{EQ} = \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_n \tag{8-1}$$

and for admittances in parallel,

$$\mathbf{Y}_{EQ} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n \tag{8-2}$$

Dividers

The useful voltage divider and its dual the current divider become, for ac circuits,

$$\mathbf{V}_2 = \frac{\mathbf{Z}_2}{\mathbf{Z}_1 + \mathbf{Z}_2}\mathbf{V} = \frac{\mathbf{Z}_2}{\mathbf{Z}_{EQ}}\mathbf{V} \quad \text{and} \quad \mathbf{I}_2 = \frac{\mathbf{Y}_2}{\mathbf{Y}_1 + \mathbf{Y}_2}\mathbf{I} = \frac{\mathbf{Y}_2}{\mathbf{Y}_{EQ}}\mathbf{I} \tag{8-3}$$

When a particular voltage or current is desired, using these relations is preferable to writing and solving loop or node equations.

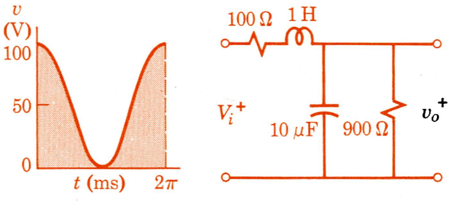
Linearity and Superposition

The concept of linearity and the theorem of superposition are inseparable; the first is essential to the second, and the second defines the first. We say that function f is linear in x if

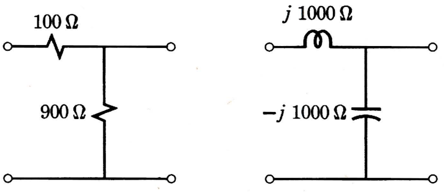
$$f(kx) = kf(x) \quad \text{and} \quad f(x_1 + x_2) = f(x_1) + f(x_2) \tag{8-4}$$

Example 1

The signal shown in Fig. 8.1a is applied to a filter circuit. Predict the output.



(a) Signal (b) Circuit



(c) DC circuit (d) AC circuit

Figure 8.1 Superposition application.

The signal can be described as

$$v_i = 50 + 50 \cos 1000t = V_i + V_{mi} \cos 1000t \text{ V}$$

For the dc component, the output is

$$V_o = \frac{R_2}{R_1 + R_2}V_i = \frac{900}{100 + 900}50 = 45 \text{ V}$$

For the ac component the resistances are negligible in comparison to the reactances; therefore, the output amplitude is

$$V_{mo} \cong \frac{Z_C}{Z_L + Z_C}V_{mi} = \frac{-j100}{j1000 - j100}50 = -5.55 \text{ V}$$

Using superposition, the total output is

$$v_o = 45 - 5.55 \cos 1000t \text{ V}$$

These defining equations paraphrase the superposition theorem (p. 48), which is particularly useful in predicting the response of a network to a complex signal. (See Example 1 on p. 210.)

Active Networks

Active networks containing ac energy sources can be represented by simpler equivalents using Thévenin's theorem or Norton's theorem.

> *Insofar as a load is concerned, any one-port network of linear elements and energy sources can be replaced by a series combination of an ideal voltage source and a linear impedance or a parallel combination of an ideal current source and a linear admittance.*

Here the Thévenin equivalent is a phasor voltage $\mathbf{V}_T$ in series with a complex impedance $\mathbf{Z}_T$; the Norton equivalent is a phasor current $\mathbf{I}_N$ in parallel with a complex admittance $\mathbf{Y}_N$. At the convenience of the circuit analyst, the equivalents are interchangeable.

Example 2

Predict the current **I** in Fig. 8.2a in response to a voltage $\mathbf{V} = 10\underline{/0^\circ}$ V.

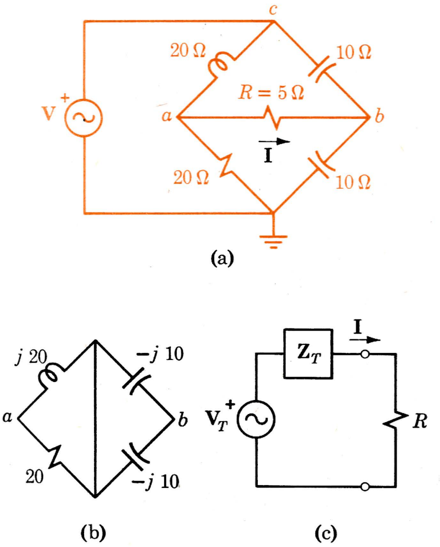


Figure 8.2 Thévenin's theorem application in an ac circuit.

One approach is to replace all except the 5-Ω resistance by a Thévenin equivalent. With R removed, $\mathbf{V}_a$ and $\mathbf{V}_b$ can be determined by voltage division and then $\mathbf{V}_T$ by subtraction.

$$\mathbf{V}_a = \frac{20}{20 + j20} \cdot 10\underline{/0^\circ} = \frac{20 \times 10\underline{/0^\circ}}{20\sqrt{2}\underline{/45^\circ}} = 5 - j5$$

$$\mathbf{V}_b = \frac{-j10}{-j10 - j10} 10\underline{/0^\circ} = \frac{10 \times 10\underline{/0^\circ}}{20} = 5 + j0$$

$$\mathbf{V}_T = \mathbf{V}_{OC} = \mathbf{V}_a - \mathbf{V}_b = -j5 = 5\underline{/-90^\circ} \text{ V}$$

With the voltage source removed ($\mathbf{V} = 0$),

$$\mathbf{Z}_T = \mathbf{Z}_{ab} = \frac{20(j20)}{20 + j20} + \frac{-j10}{2}$$

$$= 10 + j10 - j5 = 10 + j5\,\Omega$$

Then

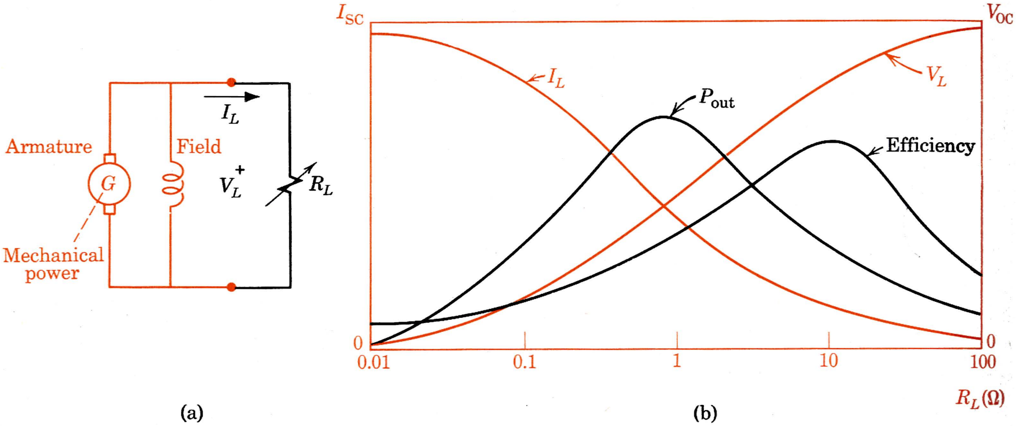
$$\mathbf{I} = \frac{\mathbf{V}_T}{R + \mathbf{Z}_T} = \frac{5\underline{/-90^\circ}}{5 + 10 + j5} = 0.32\underline{/-108.4^\circ} \text{ A}$$

Maximum Power Transfer

One of the distinguishing characteristics of the engineer is a concern with optimization. The engineer solves real problems and his or her solutions are always compromises. Efficiency costs money; safety increases complexity; performance adds weight; improvement takes time. In striving for the optimum design, the engineer provides the highest efficiency per dollar, the most powerful performance per pound, or the best results within the deadline. In addition, he must cope with the fact that the cheapest energy is in short supply and the most efficient device may be environmentally unacceptable.

Frequently the engineer has the problem of "matching" one component of a system to another to obtain optimum results. For example, an automobile transmission must be matched to the engine and, in another sense, the passing gear provides a match between the engine and the load represented by acceleration up a steep hill. The bicycle-gear ratio represents a simple solution to the problem of matching a load to a source. Why does a racing bicycle have 10 gear ratios?

The optimum load for an electric source depends on the desired result. A dc generator (Fig. 8.3a) has certain *fixed losses* that are present whether or not a load is connected and *variable losses* that are directly related to the output current. As shown in the graph, for R_L small, the output current I_L is high, but the output voltage V_L is low; for R_L large, the output voltage is high, but the current is low. At an intermediate value of R_L the power output is a maximum. The efficiency, output/input, increases as the load resistance increases from zero. Because of the fixed losses, however, a maximum efficiency occurs at a particular value of R_L. If the available power is large and efficiency is the important

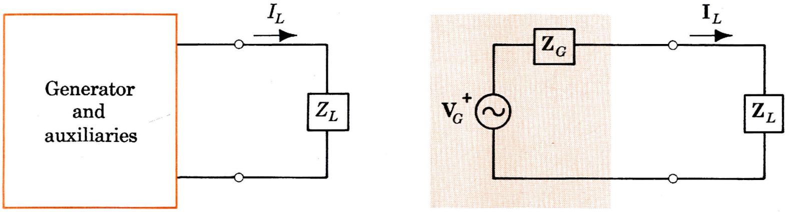


(a) (b)

Figure 8.3 A dc generator with variable load resistance.

criterion, a relatively large value of load resistance is optimum; this is true in most applications of electrical power equipment. If efficiency is less important and the power available is limited, then a lower value of load resistance is optimum; this is true in many communication applications.

With the aid of Thévenin's theorem we can draw a general conclusion about the conditions for maximum power transfer. We know that any network of linear elements and energy sources (and, approximately, any real generator and its associated circuitry) can be represented by a series combination of an ideal voltage **V** and an impedance **Z.** In the simplest case, these are the open-circuit generator voltage $\mathbf{V}_G$ and the internal impedance $\mathbf{Z}_G$ (Fig. 8.4). The current delivered to the load is



(a) Actual system (b) Linear model

Figure 8.4 Power transfer analysis.

$$I_L = \frac{V_G}{|\mathbf{Z}_G + \mathbf{Z}_L|} = \frac{V_G}{\sqrt{(R_G + R_L)^2 + (X_G + X_L)^2}}$$

The power transferred to the load is

$$P_L = I_L{}^2 R_L = \frac{V_G{}^2 R_L}{(R_G + R_L)^2 + (X_G + X_L)^2} \tag{8-5}$$

For a given generator, V_G, R_G, and X_G are fixed; R_L and X_L can be adjusted for optimum results. For maximum power transfer, Eq. 8-5 is to be maximized. The general procedure is to differentiate with respect to the variable and set the derivative equal to zero, but Eq. 8-5 indicates that for any value of R_L, P_L is maximum for $X_L = -X_G$. Hence, the first requirement for maximum power transfer is that the reactance of the load be made equal and opposite to the equivalent reactance of the source.

Under this condition,

$$P_L = \frac{V_G{}^2 R_L}{(R_G + R_L)^2} = V_G{}^2 R_L (R_G + R_L)^{-2}$$

and for maximum P_L,

$$\frac{dP_L}{dR_L} = V_G^2[(R_G + R_L)^{-2} - 2R_L(R_G + R_L)^{-3}] = 0$$

Solving,

$$2R_L = R_G + R_L \qquad \text{or} \qquad R_L = R_G$$

Hence, the second requirement is that the resistance of the load be made equal to the equivalent resistance of the source. In general, for maximum power transfer the impedance of the load should be adjusted so that

$$\mathbf{Z}_L = R_L + jX_L = R_G - jX_G = \mathbf{Z}_G^* \tag{8-6}$$

This principle can be expressed in words as follows:

> *For maximum power transfer the impedance of the load should be made equal to the complex conjugate of the Thévenin equivalent impedance of the source.*

Example 3

A generator represented by a voltage $V_G = 100\underline{/0^\circ}$ V and an internal impedance $\mathbf{Z}_G = 10 + j20\ \Omega$ is connected to a load $\mathbf{Z}_L = 20 + jX_L$ by means of a transmission network $\mathbf{Z}_T = R_T - j15\ \Omega$. Specify the values of R_T and X_L for maximum power transfer to the load.

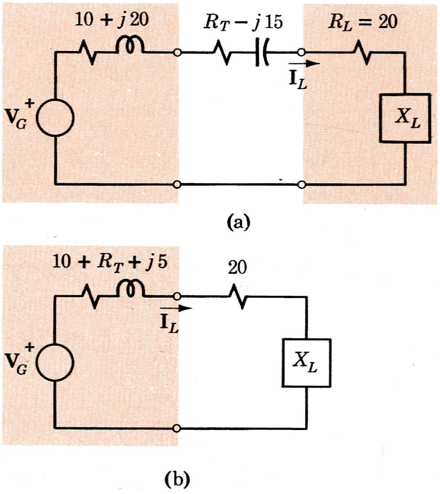


Figure 8.5 Maximizing power.

From the viewpoint of the load, the "source" consists of the generator and the transmission line. In this general case (Fig. 8.5), the "source" has one variable element R_T, and the load has one variable element X_L.

If P_L is to be maximized, then I_L must be maximized and R_T must be *zero* because the addition of any resistance can only decrease I_L.

Insofar as the load is concerned, the equivalent source impedance is then

$$\mathbf{Z}_{\text{EQ}} = \mathbf{Z}_G + \mathbf{Z}_T = 10 + j20 + 0 - j15 = 10 + j5\ \Omega$$

By Eq. 8-11, for maximum I_L and maximum P_L,

$$X_L = -X_{\text{EQ}} = -5\ \Omega$$

If a variation in R_L were permitted, then, for maximum power transfer from this source, R_L would be made equal to $R_{\text{EQ}} = 10\ \Omega$.

Under the conditions for maximum power transfer, the efficiency is just 50% since $I^2R_L = I^2R_G$ and half the power is dissipated internally in the source. No electrical appliance is designed to draw maximum power from a wall-outlet, which typically has an equivalent internal impedance of a small fraction of an ohm. Instead, an electrical load is usually designed to draw a definite amount of power and, for high efficiency and low voltage drop, the impedance of the source is made as low as is economically feasible. Note that the source impedance is *not* made equal to the load impedance; increasing the source resistance can only decrease the power transfer to a given load.

Reciprocity Theorem

Another interesting principle (which we need for a development later in this chapter) can be demonstrated by the circuit of Fig. 8.6a. For convenience in analysis, the circuit elements are replaced (Fig. 8.6b) by their equivalent impedances evaluated at the frequency of the voltage source **V**. Using the current-divider concept, the current in branch 2 is

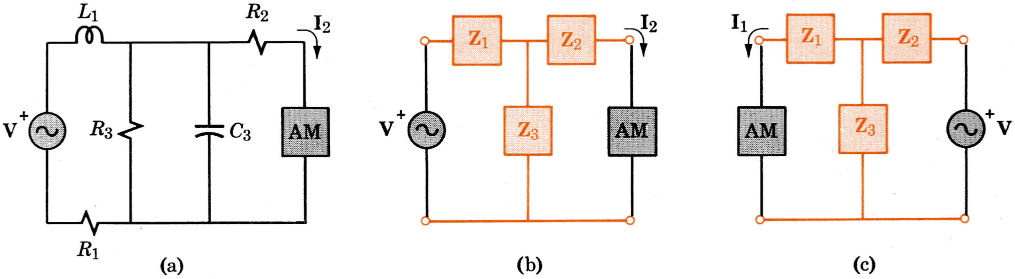


Figure 8.6 Reciprocity in linear networks.

$$\mathbf{I}_2 = \frac{\mathbf{Z}_3}{\mathbf{Z}_2 + \mathbf{Z}_3} \cdot \frac{\mathbf{V}}{\mathbf{Z}_1 + \dfrac{\mathbf{Z}_2\mathbf{Z}_3}{\mathbf{Z}_2 + \mathbf{Z}_3}} = \frac{\mathbf{Z}_3}{\mathbf{Z}_1\mathbf{Z}_2 + \mathbf{Z}_1\mathbf{Z}_3 + \mathbf{Z}_2\mathbf{Z}_3}\mathbf{V} \tag{8-7}$$

Suppose the voltage source and the ammeter are interchanged as in Fig. 8.6c. How will the current indicated by the ammeter compare to that in the original circuit? Using the current-divider concept again,

$$\mathbf{I}_1 = \frac{\mathbf{Z}_3}{\mathbf{Z}_1 + \mathbf{Z}_3} \cdot \frac{\mathbf{V}}{\mathbf{Z}_2 + \dfrac{\mathbf{Z}_1\mathbf{Z}_3}{\mathbf{Z}_1 + \mathbf{Z}_3}} = \frac{\mathbf{Z}_3}{\mathbf{Z}_1\mathbf{Z}_2 + \mathbf{Z}_2\mathbf{Z}_3 + \mathbf{Z}_1\mathbf{Z}_3}\mathbf{V} \tag{8-8}$$

The current is exactly the same! This calculation illustrates a principle, called the *reciprocity theorem*, which can be proved in general form. One statement of the reciprocity theorem is as follows:

> *In any passive, linear network, if a voltage* **V** *applied in branch 1 causes a current* **I** *to flow in branch 2, then voltage* **V** *applied in branch 2 will cause current* **I** *to flow in branch 1.*

The ratio of a voltage $\mathbf{V}_1$ in one part of a network to a current $\mathbf{I}_2$ in another part is called the *transfer impedance* $\mathbf{Z}_{12} = \mathbf{V}_1/\mathbf{I}_2$. This network parameter has all the properties of an impedance; for example, the pole-zero diagram of the transfer impedance $\mathbf{Z}_{12}$ determines the character of the response i_2 to an excitation v_1. (In a similar way, the frequency response of an amplifier is determined by a *transfer function* relating output to input variables.) An important conclusion of the reciprocity theorem is:

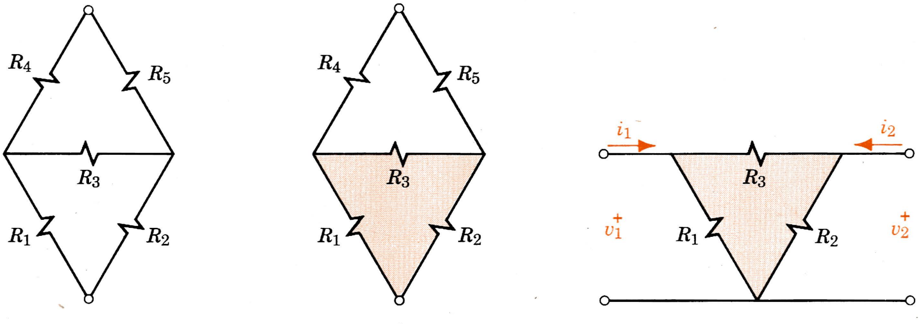
> *In passive, linear networks, the transfer impedance* $\mathbf{Z}_{12}$ *is just equal to* $\mathbf{Z}_{21}$.

In Fig. 8.6, the transfer impedance is

$$\mathbf{Z}_{12} = \frac{\mathbf{V}_1}{\mathbf{I}_2} = \mathbf{Z}_{21} = \frac{\mathbf{V}_2}{\mathbf{I}_1} = \frac{\mathbf{Z}_1\mathbf{Z}_2 + \mathbf{Z}_2\mathbf{Z}_3 + \mathbf{Z}_3\mathbf{Z}_1}{\mathbf{Z}_3} \tag{8-9}$$

Two-Port Networks

If we attempt to find the equivalent series resistance of the circuit in Fig. 8.7a by network reduction methods, we find that there are no resistances in series or parallel to be combined. There is something different about this circuit. (What is it?) The network reduction formulas previously derived are applicable to two-terminal networks only; in this circuit there is a *three-terminal network* that cannot be replaced by a two-terminal equivalent. However, if the three-terminal network shaded in Fig. 8.7b is transformed into a properly selected three-terminal equivalent network, the difficulty is removed and the equivalent series resistance can be found.



(a) (b) (c)

Figure 8.7 Analysis of a bridge circuit.

Two-Ports

The three-terminal network is a special case of the general *two-port* suggested by Fig. 8.7c. By convention, $+i_1$ and $+i_2$ are assumed *into* the network. Note that the current at the third terminal is defined by i_1 and i_2, and the third voltage is defined by v_1 and v_2. In general, of the four variables only two are independent and we may write

$$\begin{cases} v_1 = f_1(i_1, i_2) \\ v_2 = f_2(i_1, i_2) \end{cases} \quad \text{or} \quad \begin{cases} i_1 = \phi_1(v_1, v_2) \\ i_2 = \phi_2(v_1, v_2) \end{cases} \tag{8-10}$$

Equations similar to Eqs. 8-10 are useful in characterizing real devices such as transistors and motors. For the special case of ideal elements, the functions are linear and Eqs. 8-10 become

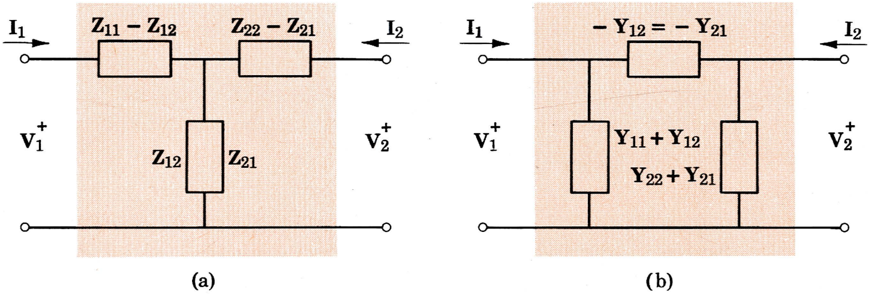
$$\begin{cases} v_1 = k_1 i_1 + k_2 i_2 \\ v_2 = k_3 i_1 + k_4 i_2 \end{cases} \quad \text{or} \quad \begin{cases} i_1 = k_5 v_1 + k_6 v_2 \\ i_2 = k_7 v_1 + k_8 v_2 \end{cases} \tag{8-11}$$

For sinusoidal excitation, the voltages and currents are represented by phasors, the constants are complex impedances or admittances, and Eqs. 8-13 become

$$\begin{cases} \mathbf{V}_1 = \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 \\ \mathbf{V}_2 = \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 \end{cases} \quad \text{or} \quad \begin{cases} \mathbf{I}_1 = \mathbf{Y}_{11}\mathbf{V}_1 + \mathbf{Y}_{12}\mathbf{V}_2 \\ \mathbf{I}_2 = \mathbf{Y}_{21}\mathbf{V}_1 + \mathbf{Y}_{22}\mathbf{V}_2 \end{cases} \tag{8-12}$$

For linear elements, $\mathbf{Z}_{12} = \mathbf{Z}_{21}$ and $\mathbf{Y}_{12} = \mathbf{Y}_{21}$ by the theorem of reciprocity, and we conclude that any linear two-port can be characterized by a set of three parameters.

Comparing Eq. 8-12a with the circuit of Fig. 8.8a, we see that one possible arrangement of the impedance parameters is in a T configuration as shown. Writing loop equations for Fig. 8.8a,



(a) (b)

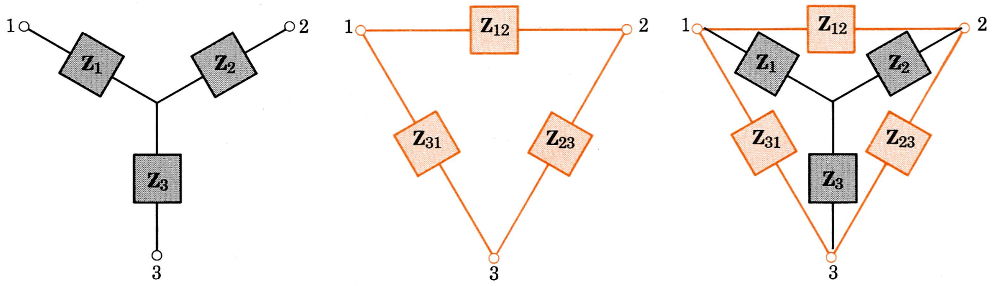
Figure 8.8 Equivalent T and Π networks.

$$\begin{aligned} \mathbf{V}_1 &= (\mathbf{Z}_{11} - \mathbf{Z}_{12} + \mathbf{Z}_{12})\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 = \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 \\ \mathbf{V}_2 &= \mathbf{Z}_{21}\mathbf{I}_1 + (\mathbf{Z}_{22} - \mathbf{Z}_{21} + \mathbf{Z}_{21})\mathbf{I}_2 = \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 \end{aligned}$$

as required. Comparing Eq. 8-12b with the circuit of Fig. 8.8b, we see that one possible arrangement of the admittance parameters is in a **Π** configuration as shown. This can be confirmed by writing the appropriate node equations. Other circuit arrangements are possible and will be developed when needed.

T-Π *Transformation*

For our present purposes, let us consider the **T** and **Π** configurations as three-terminal networks. As redrawn in Fig. 8.9, these can also be called **Y** and **Δ** networks. Here all elements are designated by their impedances (rather than admittances) and lettered in a consistent scheme.



(a) **T** or **Y** (b) **Π** or **Δ** (c) Equivalents

Figure 8.9 Equivalent three-terminal networks.

We specified that two passive two-terminal networks are equivalent if they have the same terminal impedance; if this is true, one can be substituted for the other. Extending this concept of equivalence to three-terminal networks, we say that the **T** and **Π** networks are equivalent if the impedance at any pair of terminals of one network is just equal to the impedance at the corresponding pair of terminals of the other. On the basis of this equality we can determine the conditions under which one network can be substituted for the other. As noted previously, in practical networks the equivalence may hold for only a single frequency.

The procedure is straightforward, but the algebra is laborious. Equating the impedances of the two networks between terminals 1 and 2:

$$\mathbf{Z}_1 + \mathbf{Z}_2 = \frac{\mathbf{Z}_{12}(\mathbf{Z}_{23} + \mathbf{Z}_{31})}{\mathbf{Z}_{12} + \mathbf{Z}_{23} + \mathbf{Z}_{31}} \tag{8-13}$$

Between terminals 2 and 3:

$$\mathbf{Z}_2 + \mathbf{Z}_3 = \frac{\mathbf{Z}_{23}(\mathbf{Z}_{31} + \mathbf{Z}_{12})}{\mathbf{Z}_{12} + \mathbf{Z}_{23} + \mathbf{Z}_{31}} \tag{8-14}$$

Between terminals 3 and 1:

$$\mathbf{Z}_3 + \mathbf{Z}_1 = \frac{\mathbf{Z}_{31}(\mathbf{Z}_{12} + \mathbf{Z}_{23})}{\mathbf{Z}_{12} + \mathbf{Z}_{23} + \mathbf{Z}_{31}} \tag{8-15}$$

Note the symmetry of these three equations; it is due to the inherent symmetry of the networks. The second equation can be written directly from the first by a cyclical change in subscripts.

To solve for the T (or Y elements in terms of the Π (or Δ) elements, we first eliminate $\mathbf{Z}_3$ by subtracting Eq. 8-15 from Eq. 8-14. Then this result is subtracted from Eq. 8-13 to eliminate $\mathbf{Z}_2$. The result is:

$$\mathbf{Z}_1 = \frac{\mathbf{Z}_{12}\mathbf{Z}_{31}}{\mathbf{Z}_{12} + \mathbf{Z}_{23} + \mathbf{Z}_{31}} \tag{8-16}$$

By symmetry we conclude that the impedance of the equivalent Y element is the product of the adjacent Δ elements divided by the sum of the Δ elements (Fig. 8.9c). This is called the **Δ**-Y (delta-wye) or **Π**-T (pi-tee) transformation.

If instead we solve for the equivalent Δ (or Π) elements in terms of the Y (or T) elements, the result is:

$$\mathbf{Z}_{12} = \frac{\mathbf{Z}_1\mathbf{Z}_2 + \mathbf{Z}_2\mathbf{Z}_3 + \mathbf{Z}_3\mathbf{Z}_1}{\mathbf{Z}_3} \tag{8-17}$$

Again we invoke symmetry and conclude that the impedance of the equivalent Δ element is the sum of the products of the Y elements divided by the opposite Y element.

Example 4

Three equal impedances $\mathbf{Z} = 10\underline{/60^\circ}\ \Omega$ are connected in Y to form a three-phase load (Fig. 8.10). Determine the impedances of the equivalent **Δ**.

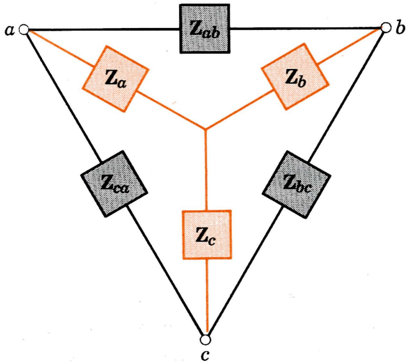


Figure 8.10 Y to **Δ** conversion.

For this balanced network, Eq. 8-17 yields

$$\mathbf{Z}_\Delta = \mathbf{Z}_{ab} = \frac{\mathbf{Z}_a\mathbf{Z}_b + \mathbf{Z}_b\mathbf{Z}_c + \mathbf{Z}_c\mathbf{Z}_a}{\mathbf{Z}_c}$$

$$= \frac{3(\mathbf{Z}_Y)^2}{\mathbf{Z}_Y} = 3\mathbf{Z}_Y = 30\underline{/60^\circ}\ \Omega$$

The impedances of the equivalent **Δ** load are just equal to three times the corresponding impedances of the balanced Y load.

Another formulation of Eq. 8-17 is obtained if we recognize and take advantage of the duality relation between the **Y** and **Δ** networks. From Eq. 8-16 we can write directly

$$\mathbf{Y}_{12} = \frac{\mathbf{Y}_1\mathbf{Y}_2}{\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3} \tag{8-18}$$

These are not equations to be memorized; rather they are formulas to be applied in specific problems. A common application is in the simplification of networks where two-terminal network reduction methods do not work.

Example 5

Determine the equivalent series resistance of the bridge network of Fig. 8.11a.

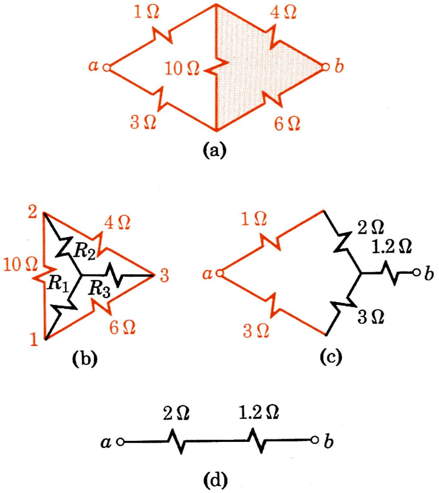


Figure 8.11 Y-**Δ** transformation.

First, the shaded **Δ** is replaced by an equivalent **Y** (Fig. 8.11b), using Eq. 8.16.

$$\mathrm{Z}_1 = \frac{\mathrm{Z}_{12}\mathrm{Z}_{31}}{\mathrm{Z}_{12} + \mathrm{Z}_{23} + \mathrm{Z}_{31}} = \frac{10 \times 6}{10 + 6 + 4} = \frac{60}{20} = 3\ \Omega$$

$$\mathrm{Z}_2 = \frac{\mathrm{Z}_{23}\mathrm{Z}_{12}}{\mathrm{Z}_{12} + \mathrm{Z}_{23} + \mathrm{Z}_{31}} = \frac{4 \times 10}{10 + 6 + 4} = \frac{40}{20} = 2\ \Omega$$

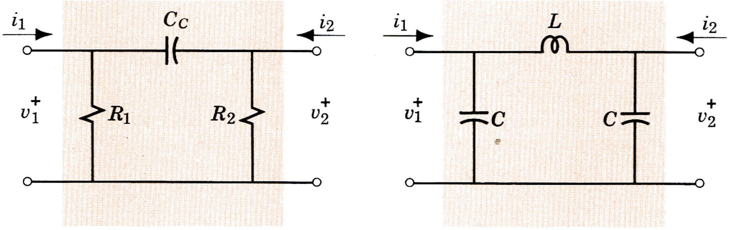
$$\mathrm{Z}_3 = \frac{\mathrm{Z}_{31}\mathrm{Z}_{23}}{\mathrm{Z}_{12} + \mathrm{Z}_{23} + \mathrm{Z}_{31}} = \frac{6 \times 4}{10 + 6 + 4} = \frac{24}{20} = 1.2\ \Omega$$

When this substitution is made, resistances in series and parallel can be combined to give

$$\begin{aligned} \mathrm{Z}_{ab} &= \frac{(3 + 3)(1 + 2)}{(3 + 3) + (1 + 2)} + 1.2 \\ &= \frac{6 \times 3}{6 + 3} + 1.2 = 2 + 1.2 = 3.2\ \Omega \end{aligned}$$

Coupling Circuits

The two-port is frequently used as a *coupling circuit* to tie one component of a system to another in an optimum way. The **Π** circuit of Fig. 8.12a is used to couple two stages of amplification; the input voltage v_1 contains ac signals and a dc component inherent in the operation of the amplifier (a vacuum tube



(a) *RC* coupling (b) Filter

Figure 8.12 Two-ports as coupling circuits.

or transistor). The output voltage contains only ac signals because the coupling capacitor C_C blocks the dc component.

The inverse function is performed by the *filter* (Fig. 8.12b), a special coupling circuit that blocks ac components but permits direct currents to flow unimpeded. Electronic power supplies that transform ac power into direct currents to supply electronic apparatus ordinarily contain filters of this type. The frequency response of these coupling circuits is interesting and will be investigated later on.

The two-ports examined so far provide *conductive* coupling, so called because there is a direct electrical connection between the output and the input. Another possibility is to employ a magnetic field to couple energy from one component to another; in this important category are transformers and induction motors. In the transformer, the output is electrically isolated from the input. In the induction motor, the rotating winding is mechanically and electrically isolated from the stationary winding. From a circuit viewpoint, these important practical features are described in terms of the circuit parameter called "mutual inductance."

Mutual Inductance

As you may recall from physics, the voltage v in volts induced in a coil of n turns encircling a changing magnetic flux ϕ in webers is

$$v = n\frac{d\phi}{dt} \tag{8-19}$$

If the flux is due to current flowing in the coil itself (Fig. 8.13a), the flux is proportional to the current and the number of turns or $\phi = kni$ and

$$v = n\frac{d(kni)}{dt} = kn^2\frac{di}{dt} = L\frac{di}{dt} \tag{8-20}$$

where L is the *self-inductance* in henrys. Up to this point, only this form of inductance has been discussed, so we have referred to L simply as "inductance."

If the changing flux is due to a current i_2 in a second coil (Fig. 8.13b), the

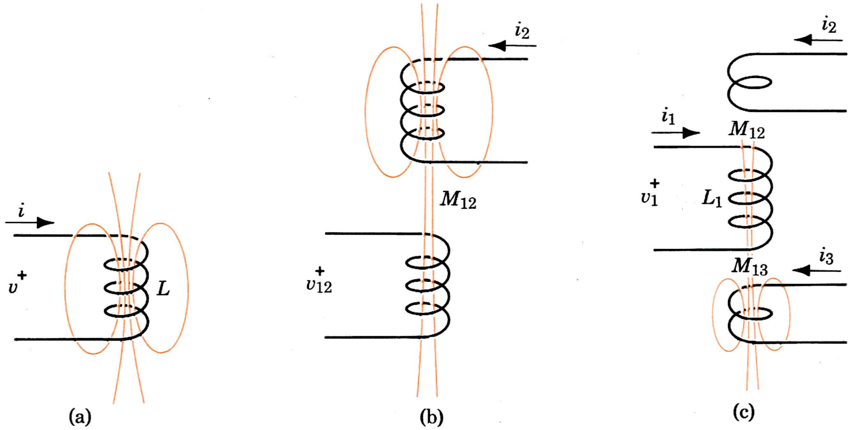


Figure 8.13 Examples of self and mutual inductance.

voltage v_{12} induced in coil 1 is

$$v_{12} = M_{12}\frac{di_2}{dt} \tag{8-21}$$

where M_{12} is the *mutual inductance* in henrys. In the same way, a voltage can be induced in coil 1 due to a changing current i_3 in a third coil. For a linear system, superposition applies and the voltage appearing across coil 1 in Fig. 8.13c is

$$v_1 = v_{11} + v_{12} + v_{13} = L_1\frac{di_1}{dt} + M_{12}\frac{di_2}{dt} + M_{13}\frac{di_3}{dt} \tag{8-22}$$

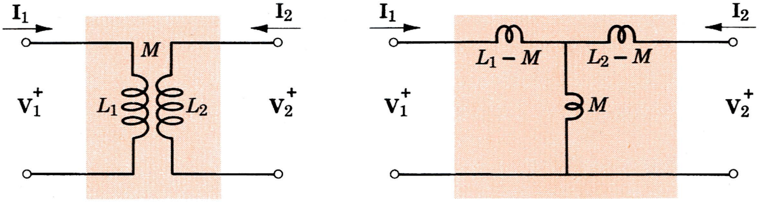
The magnitude and sign of mutual inductance M_{12} depend on the proximity and the orientation of the two coils. If the orientation is such that the flux contributed by current i_2 subtracts from that due to current i_1, the corresponding voltage term v_{12} is negative. If it is desired that mutual inductance M_{12} always be a positive quantity, then a negative sign must be used with the $M_{12}(di_2/dt)$ term in Eq. 8-22. There are conventional methods of designating coil terminals so that this sign is determined automatically.†

Inductive Coupling

A practical form of magnetic or inductive coupling is shown in Fig. 8.14a. To increase the coupling, that is, to increase the mutual inductance M, the two real coils represented by this linear model may be wound on a common core with favorable magnetic properties, but the two coils are electrically isolated. This inductively coupled two-port appears in a wide variety of practical forms. In a radio receiver, a few turns of fine wire on an insulating form couple the antenna to the first amplifier stage. In a power system, many turns of heavy wire

---

† See p. 332 of H. H. Skilling: *Electrical Engineering Circuits,* 2nd ed., John Wiley & Sons, New York, 1965.



(a) (b)

Figure 8.14 An inductively coupled circuit and the equivalent T circuit.

wound on an iron core 3 meters tall couple the generator to the high-voltage transmission line.

For purposes of analysis it may be desirable to replace the inductively coupled two-port by the equivalent T circuit of Fig. 8.14b. For sinusoids, the voltage developed in the input by the input current is $\mathbf{V}_{11} = j\omega L_1 \mathbf{I}_1$ where $j\omega L_1$ is the self-impedance $\mathbf{Z}_{11}$. The voltage $\mathbf{V}_{12}$ coupled into the input by a current $\mathbf{I}_2$ is $\mathbf{V}_{12} = j\omega M_{12}\mathbf{I}_2$ where $j\omega M_{12}$ is a mutual impedance $\mathbf{Z}_{12}$. By the reciprocity theorem, for linear networks $\mathbf{Z}_{12} = \mathbf{Z}_{21}$ or $M_{12} = M_{21} = M$ as shown. Then

$$\begin{aligned} \mathbf{V}_1 &= j\omega L_1 \mathbf{I}_1 + j\omega M \mathbf{I}_2 = \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 \\ \mathbf{V}_2 &= j\omega M \mathbf{I}_1 + j\omega L_2 \mathbf{I}_2 = \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 \end{aligned} \tag{8-23}$$

Comparing Eqs. 8-23 with the circuit of Fig. 8.14b, we see that the T circuit shown is equivalent. In the equivalent circuit, the common element is a self-inductance of magnitude M.

Example 6

A series RLC circuit is inductively coupled to an input circuit consisting of R_1 and L_1 (Fig. 8.15). Determine the total input impedance.

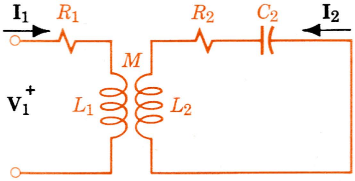


Figure 8.15 Coupled impedance.

The "primary" voltage is the sum of the $I_1 Z_{11}$ voltage drop and the voltage induced by the "secondary" current or

$$\mathbf{V}_1 = \mathbf{I}_1\mathbf{Z}_{11} + j\omega M \mathbf{I}_2 = \mathbf{I}_1(R_1 + j\omega L_1) + j\omega M \mathbf{I}_2$$

For the secondary loop, self-impedance $\mathbf{Z}_{22}$ consists of L_2, R_2, and C_2. Therefore,

$$\begin{aligned} \mathbf{V}_2 &= j\omega M \mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 = 0 \\ &= j\omega M \mathbf{I}_1 + \left(R_2 + j\omega L_2 - j\frac{1}{\omega C_2}\right)\mathbf{I}_2 \end{aligned} \tag{8-24}$$

Solving Eq. 8-24 for $\mathbf{I}_2$ and substituting,

$$\mathbf{V}_1 = \mathbf{I}_1\mathbf{Z}_{11} + j\omega M\left(-\frac{j\omega M \mathbf{I}_1}{\mathbf{Z}_{22}}\right)$$

Then the input impedance is

$$\mathbf{Z}_{\text{in}} = \frac{\mathbf{V}_1}{\mathbf{I}_1} = \mathbf{Z}_{11} + \frac{(\omega M)^2}{\mathbf{Z}_{22}} \tag{8-25}$$

In Example 6, the effect of the secondary circuit on the input impedance is as if an impedance $(\omega M)^2/\mathbf{Z}_{22}$ were connected in series with the primary. At resonance, $\mathbf{Z}_{22}$ is a relatively small pure resistance, but the impedance coupled into the primary may be a relatively large resistance.

Transformer Coupling

If the two coils of a transformer† are wound on the same highly permeable iron core, both coils are linked by essentially the same magnetic flux ϕ. Since $v = N\,d\phi/dt$, a changing flux in the core induces voltages in the windings proportional to the number of turns. In the ideal transformer of Fig. 8.16, (closely approximated by a real transformer),

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \tag{8-26}$$

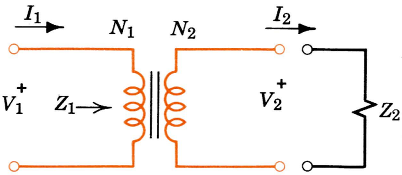


Figure 8.16 An ideal transformer.

The current ratio can be predicted by noting that in an ideal transformer there is no power loss, and the power in must equal the power out. Therefore, a *step-down* in current must accompany a *step-up* in voltage or

$$\frac{I_2}{I_1} = \frac{V_1}{V_2} = \frac{N_1}{N_2} \tag{8-27}$$

If an impedance Z_2 is connected to the *secondary*, the impedance seen at the *primary* is, by Eqs. 8-26 and 8-27,

$$Z_1 = \frac{V_1}{I_1} = \frac{(N_1/N_2)V_2}{(N_2/N_1)I_2} = \left(\frac{N_1}{N_2}\right)^2 \frac{V_2}{I_2} = \left(\frac{N_1}{N_2}\right)^2 Z_2 \tag{8-28}$$

On the basis of this brief discussion, we can see three important advantages of the transformer as a coupling device:

1. AC power is easily and efficiently transformed from the optimum generating voltage of, say, 25 kV up to an efficient long-distance transmission voltage of 500 kV, down to an optimum urban distribution voltage of 50 kV, and then down to a safe utilization voltage of 120 or 240 V.
2. Only a changing flux induces voltage; direct currents are not "transformed." In an amplifier, for example, a transformer passes on the ac signal but filters out dc currents that might adversely affect some amplifier components.
3. The transformer transforms impedance and facilitates impedance matching. By proper choice of the turns ratio, the impedance of a given load can be made to appear to be optimum for a given source. For example, the low resistance of a loud-speaker can be matched to the much higher output resistance of an amplifier.

---

† For a detailed discussion see pages 576–580.

SUMMARY

- Two passive one-ports are equivalent if they have the same input impedance $\mathbf{Z}$ or the same input admittance $\mathbf{Y}$.
  In general, two real networks are equivalent at one frequency only.
  For impedances in series, $\mathbf{Z}_{EQ} = \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_n$
  For admittances in parallel, $\mathbf{Y}_{EQ} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n$
- If cause and effect are linearly related, the principle of superposition applies and the total effect of several causes acting simultaneously is equal to the sum of the effects of the individual causes acting one at a time.
- Insofar as a load is concerned, any one-port network of linear elements and energy sources can be replaced by a series combination of an ideal voltage source $\mathbf{V}_T$ and a linear impedance $\mathbf{Z}_T$ (Thévenin's equivalent), or by a parallel combination of an ideal current source $\mathbf{I}_N$ and a linear admittance $\mathbf{Y}_N$ (Norton's equivalent), where $\mathbf{V}_T = \mathbf{V}_{\text{OC}}$, $\mathbf{Z}_T = \mathbf{V}_{\text{OC}}/\mathbf{I}_{\text{SC}}$, $\mathbf{I}_N = \mathbf{I}_{\text{SC}}$, and $\mathbf{Y}_N = \mathbf{I}_{\text{SC}}/\mathbf{V}_{\text{OC}} = 1/\mathbf{Z}_T$.
- For maximum power transfer, the impedance of the load should be made equal to the complex conjugate of the Thévenin equivalent impedance of the source.
- The reciprocity theorem says: In any passive, linear network, if a voltage $\mathbf{V}$ applied in branch 1 causes a current $\mathbf{I}$ to flow in branch 2, then voltage $\mathbf{V}$ applied in branch 2 will cause current $\mathbf{I}$ to flow in branch 1.
  In such a network, transfer impedance $\mathbf{Z}_{12} = \mathbf{V}_1/\mathbf{I}_2 = \mathbf{V}_2/\mathbf{I}_1 = \mathbf{Z}_{21}$.
- Any linear two-port network can be characterized by a set of three parameters and represented by an equivalent T (Y) or **Π** (**Δ**) network.
- Coupling circuits are two-ports used to connect one component of a system to another in an optimum way.
  If two ideal coils are inductively coupled, mutual inductance exists and the voltage across coil 1 is

$$v_1 = L_1\frac{di_1}{dt} + M\frac{di_2}{dt}$$

  Transformer coupling optimizes power transmission, filters out dc currents, and permits impedance matching. Ideally,

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \qquad \frac{I_2}{I_1} = \frac{N_1}{N_2} \qquad \frac{Z_1}{Z_2} = \left(\frac{N_1}{N_2}\right)^2$$

REVIEW QUESTIONS

1. What is the difference between a "dual" and an "equivalent"?
2. Explain in words the two equations used to define linearity.
3. Under what circumstances is Thévenin's theorem useful?
4. Is the same power dissipated in an active circuit and the Norton equivalent?
5. How are voltage and current sources "removed"?
6. Energy is to be added to a ball by throwing. For maximum energy would you choose a marble, a golf ball, a baseball, or a medicine ball? Why?
7. Why is the gear ratio different on boys' and girls' bicycles?
8. Why is maximum power transfer sometimes undesirable?
9. Explain how impedances are matched by using a transformer.
10. How is the fact that $\mathbf{Z}_{12} = \mathbf{Z}_{21}$ related to the reciprocity theorem?
11. Draw a T network of resistances and add a resistance between terminals 1 and 2. Outline the procedure for reducing this to a Π network.
12. Is the same power dissipated in a Δ three-phase load and the equivalent Y load?
13. Sketch, qualitatively, the frequency response curve of V_2/V_1 versus ω for the coupling circuit of Fig. 8.12a.
14. Why is inductive coupling advantageous in a motor?

EXERCISES

1. A one-port consists of $\mathbf{Z}_1 = 10\underline{/+30^\circ}\ \Omega$ and $\mathbf{Z}_2 = 20\underline{/-45^\circ}\ \Omega$ connected in parallel.
   (a) Find R and X of an equivalent series combination.
   (b) Find G and B of an equivalent parallel combination.
   (c) For $\mathbf{I} = 10\underline{/0^\circ}$ A into the one-port, find $\mathbf{I}_1$ and $\mathbf{I}_2$.
2. In Fig. 8.17, switch S is closed at $t = 0$. Use Thévenin's theorem to simplify the circuit and find the time constant for this circuit. Determine $i_L(t)$.

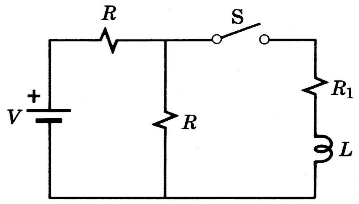


Figure 8.17

3. In Fig. 8.18, the reactances are given in ohms for $\omega = 1000$ rad/s. Find the steady-state current in L if:
   (a) $\mathbf{V}$ is replaced by $v = 120 + 120 \cos 1000t$ V.

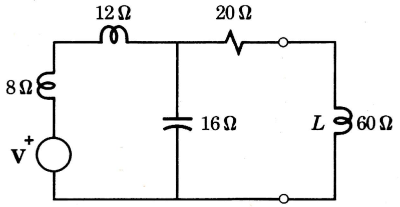


Figure 8.18

   (b) $\mathbf{V}$ is replaced by $v = 120 \cos 1000t + 120 \cos 4000t$ V.
4. In Fig. 8.18, $\mathbf{V} = 20\underline{/-90^\circ}$ V and the load consists of an inductive reactance $X_L = 60\ \Omega$. Using Norton's theorem, replace the remainder of the network by a simple combination of current source and admittance and determine the current in X_L. Check your answer using loop equations.
5. In Fig. 8.19, a variable load R_L is supplied through a transmission line $R_T = 15\ \Omega$ from a voltage generator $V = 600$ V, and a current generator $I = 10$ A.
   (a) Replace the two generators by a single equivalent current generator.
   (b) Replace the generators and transmission line by a single equivalent voltage generator.

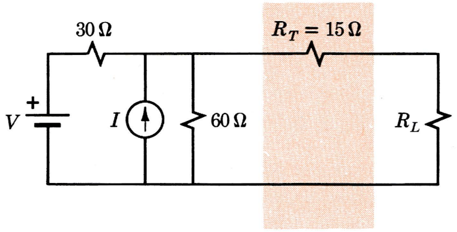


Figure 8.19

(c) Determine the power delivered to a load $R_L = 25\ \Omega$.
(e) Determine the power loss in R_T.

6. Repeat part (c) of Exercise 5 using superposition.
7. A current $\mathbf{I} = 2\underline{/0^\circ}$ A flows between terminals 1 and 2 in a complicated linear network when a voltage $\mathbf{V} = 10\underline{/90^\circ}$ V is applied between terminals 3 and 4. If a voltage $\mathbf{V} = 25\underline{/+45^\circ}$ V is applied between terminals 1 and 2, what current would be observed to flow between terminals 3 and 4? State theorems used.
8. For the two-port of Fig. 8.20, evaluate the parameters k_5, k_6, k_7, k_8 (Eq. 8-11b). Determine the parameters of an equivalent **Π** network following the scheme of Fig. 8.8b.

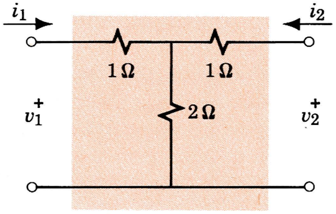


Figure 8.20

9. From Eqs. 8-13, 14, and 15, derive Eq. 8-16 for the equivalent **Y** element $\mathbf{Z}_1$.
10. Derive Eq. 8-17 for the equivalent **Δ** element $\mathbf{Z}_{12}$.
11. In Fig. 8.20, the right-hand 1-Ω resistance is removed (shorted out). For the new **T** network, derive the equivalent **Π**.
12. Simplify the circuit of Fig. 8.2a by **T-Π** transformation and determine the current in the 20-Ω resistance.
13. Simplify the circuit of Fig. 8.2a by **T-Π** transformation and determine the current in the voltage source **V**.
14. The input to the coupling circuit of Figure 8.12a is $v_1 = 10 + 10 \cos 10t + 10 \cos 1000t$. If $R_1 = R_2 = 10\ \text{k}\Omega$ and $C_C = 1\ \mu\text{F}$, predict v_2. In what sense is this a *filter*?
15. The input to the coupling circuit of Figure 8.12b is $v_1 = 10 + 10 \cos 500t + 10 \cos 2000t$. If $C = 10\ \mu\text{F}$ and $L = 10$ H, predict v_2. In what sense is this a *filter*?
16. In Fig. 8.21, $\mathbf{I} = 6$ A and $R_L = 8\ \Omega$.
    (a) Specify voltage **V** so that $\mathbf{I}_L$ is 3.0 A.

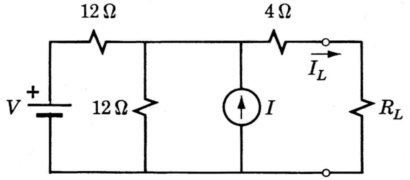


Figure 8.21

    (b) Specify the value of R_L so that a maximum power is delivered to R_L.
    (c) For $V = 48$ V and $R_L =$ the value specified in part (b), predict the power delivered to R_L.
17. In the circuit of Fig. 8.22, $i = 2 \cos 10{,}000t$ A, $R_1 = 5\ \Omega$ and $C = 10\ \mu\text{F}$. Design the load, i.e., specify values of L and R, for maximum power transfer to R and predict the value of P_{max}.

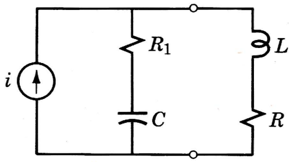


Figure 8.22

18. For the circuit of Exercise 5,
    (a) Predict the value of R_L to absorb maximum power.
    (b) Plot 3 points on a graph of P_L versus R_L and check your prediction.
19. Measurements on a human nerve cell indicate an open-circuit voltage of 80 mV, and

a current of 5 nA through a 6-MΩ load. You are to predict the behavior of this cell with respect to an external load.
(a) What simplifying assumption must be made?
(b) What is the maximum power available from the cell?

20. A load $R_L = 20\ \Omega$ is supplied through a transmission line R_T from a generator $V_G = 240$ V, $R_G = 30\ \Omega$ as shown in Fig. 8.23. Specify the value of R_T for:
(a) Maximum power dissipated in R_T.
(b) Maximum power supplied *by* the generator.
(c) Maximum power delivered *to* the load.

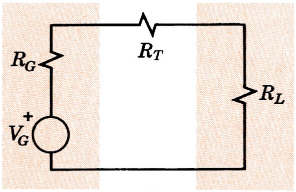


Figure 8.23

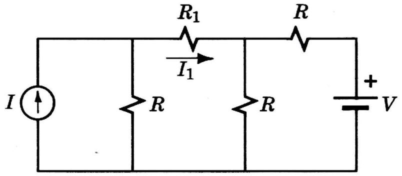


Figure 8.24

21. In Fig. 8.24, source V is replaced by a high-resistance voltmeter. $I = 12$ A, $R = 10\ \Omega$, and $R_1 = 20\ \Omega$.
(a) Predict the reading on the voltmeter.
(b) If current source I and the voltmeter are interchanged, predict the new voltmeter reading. Check analytically.

22. When two coils are connected in series, the total inductance is measured to be 50 mH. When the connections to one coil are reversed, the total series inductance is 30 mH.
(a) What is the mutual inductance?
(b) If the coils are similar, except that coil 2 has twice as many turns as coil 1, what are L_1 and L_2?

23. For the circuit of Fig. 8.15, sketch the magnitude of the input impedance as a function of frequency near the series resonant frequency of the secondary.

24. In Fig. 8.14a, $L_1 = 1$ H, $L_2 = 2$ H, and $M = 0.5$ H. A capacitance $C = 8\ \mu$F is connected across the output and a current $\mathbf{I}_1 = 15\underline{/0^\circ}$ A at $\omega = 500$ rad/s is applied. Determine the voltage $\mathbf{V}_2$ and the input impedance $\mathbf{Z}_1$.

25. An ideal transformer (with 5 turns in the primary winding for every turn in the secondary) can deliver continuously a 10-A current at 250 V to a load. If the load is a pure resistance, determine:
(a) The primary voltage and current.
(b) The power input to the primary.
(c) The impedance seen at the primary terminals.

PROBLEMS

1. Telephone transmission lines can be represented by the ladder network shown in Fig. 8.25 where L and C are the inductance and capacitance per unit length. If the line is long, the impedance becomes equal to the *characteristic impedance* $\mathbf{Z}_0$, the impedance looking into an *infinite* line.
(a) If the line is infinitely long, what is the impedance of the section to the right of terminals *aa*?
(b) Use this fact to evaluate the characteristic impedance of this line.

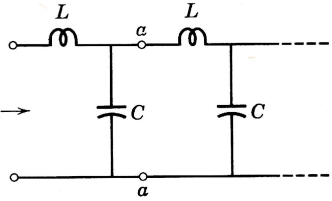


Figure 8.25

(c) Let the unit length over which L and C are measured approach zero and evaluate $\mathbf{Z}_0$.

2. In Fig. 8.26, $i_1 = 15 + 10 \cos 2000t$ A, $R_1 = 4\ \Omega$, $C = 125\ \mu\text{F}$, $L = 1$ mH, $R = 2\ \Omega$, and $v = 20 \cos 2000t$ V. Clearly indicating the methods used, take advantage of network theorems to determine the steady-state current i.

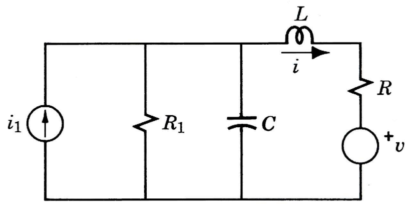


Figure 8.26

3. A telephone transmitter with an output resistance $R_o = 600\ \Omega$ is to feed a transmission line that can be represented by an infinite ladder of resistances as in Fig. 8.27. Design the transmission line, i.e., specify R for maximum power transfer.

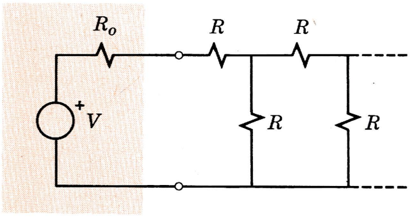


Figure 8.27

4. A load **Z** is supplied from two energy sources $\mathbf{I} = 10\underline{/90^\circ}$ A and $\mathbf{V} = 200\underline{/0^\circ}$ V as shown in Fig. 8.28. Specify the value of **Z** to absorb maximum power, and predict the power absorbed by the specified value of **Z**.

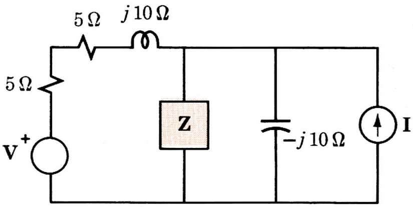


Figure 8.28

5. An ac generator has an open-circuit voltage $V = 20$ kV and an internal impedance $\mathbf{Z} = 2 + j10\ \Omega$. It is connected to a load through an underground transmission line that can be represented by a *parallel* capacitance $X_C = 12\ \Omega$. The impedance of the load can be varied in magnitude and angle. Draw the circuit and determine the maximum power that can be transmitted from the generator to the load.

6. In Fig. 8.29, load resistance R_L is equal to $R_G/2$.
   (a) Determine the power transferred to R_L if connected directly to the generator.
   (b) To increase power transfer, the coupling circuit shown is suggested as an "impedance-matching" device. Define the relation between X_1, X_2, and X_3 that must be satisfied if power transfer is to be maximum.
   (c) Outline a procedure for designing the coupling circuit.
   (d) By what factor would the power transferred be increased by proper impedance matching?

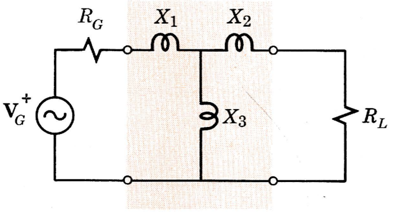


Figure 8.29

7. A coil with an inductance of 50 μH and a Q of 50 is inductively coupled ($M = 50\ \mu\text{H}$) to a second coil with $L_2 = 100\ \mu\text{H}$ and $Q = 50$ (see Fig. 8.15). The secondary is tuned to series resonance at $f_0 = 1.59$ MHz by means of C_2. Predict the input impedance Z_1 at the resonant frequency.

8. In Fig. 8.30, L_1 is fixed, but the number of turns on coil 2 may be adjusted so that L_2

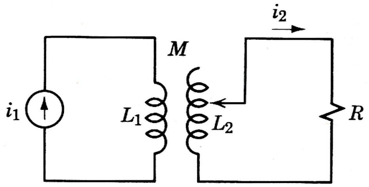


Figure 8.30

may be chosen at will. However, assume that the coefficient of coupling $k = M/\sqrt{L_1 L_2}$ will remain constant. If $i_1 = I_m \cos \omega t$, what value of L_2 will lead to the most power in R?

9. A signal generator has an internal resistance $R_G = 500\ \Omega$. Design a coupling circuit (specify values of L and C in Fig. 8.31) to provide maximum power transfer at $\omega = 10{,}000$ rad/s to a load $R_L = 5000\ \Omega$.

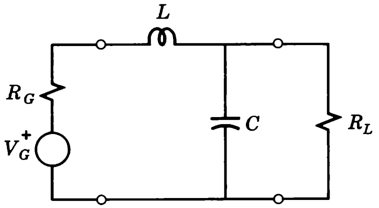


Figure 8.31

Chapter 9
Introduction to Systems

Systems Engineering
Block Diagrams
Feedback Circuits
Transfer Functions

A system is a combination of diverse but interacting elements integrated to achieve an overall objective. The elements may be devices, plants, or human beings concerned with processing materials, information, or energy. The objective may be to guide a space vehicle, to control a chemical process, or to provide a valuable service. In the health care system we call a hospital, for example, the human elements include doctors, nurses, and technicians using a great variety of instruments and apparatus, and employing information in the form of observations and records. Patients are admitted, observed, prepared, treated, charged, and discharged. Actions by the doctor produce changes in the patient that are reported back to the doctor who compares patient condition to a reference and makes a decision that is communicated to the dispensary that transmits medicine to the nurse for administering to the patient.

Systems Engineering

The increasing complexity of man-made systems and the growing body of principles and techniques for predicting system behavior have resulted in the new activity called *systems engineering.* The systems engineer takes an overall view; he is interested in specifying the performance of the system, creating a combination of components to meet the specifications, describing the characteristics of components and their interconnections, and evaluating system performance. He leaves the design of components and connecting links to others; his focus is on external characteristics—performance, reliability, and cost.

Guidance System

From the standpoint of the systems engineer, a device or a subsystem can be represented by a simple *block* (or model) labeled to indicate the function performed (without indicating *how* it is performed). In the space vehicle guidance system in Fig. 9.1, transducer T converts the desired heading θ_R into reference signal V_R, and sensor S converts the actual heading θ into signal V_F, which is fed back to the input. Error detector (or summer) Σ compares the actual heading with the desired heading and amplifier A magnifies the difference. An actuator (not shown) opens vernier jet *J*, which accelerates capsule C into the new heading. The elements and their outputs are diverse in nature, but no one element can be considered alone; the interconnections, particularly the feedback loop, determine the overall behavior. For harmonious operation and optimum performance the elements must be matched in such characteristics as impedance and frequency response.

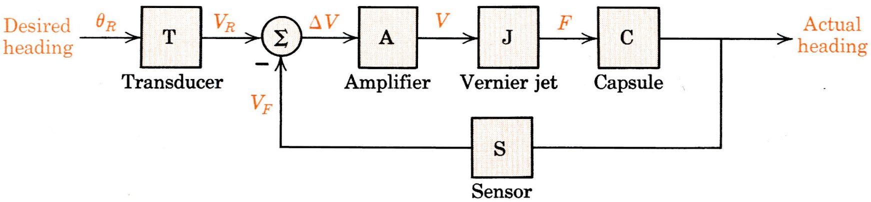


Figure 9.1 A space vehicle guidance system.

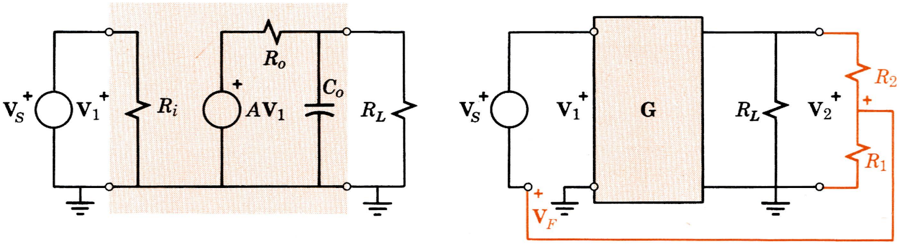
A feedback loop for returning part of the output to the input in order to improve performance is an important component in many systems. The concept of feedback is not new; it is inherent in many biological systems, and it is found in well-established social systems. However, the formulation of the principles of feedback is a recent achievement (credited to H. S. Black in his studies in 1934), and exploitation of this new understanding has come in just the last few years.

In electrical systems there are two general categories of feedback applications. One is concerned with the use of feedback for purposes of comparison and subsequent control, as in the vehicle guidance system (Fig. 9.1) and the automatic voltage regulator (Fig. 26.14). The analysis of the behavior of such control systems is the subject of Chapter 26.

A second application of feedback is to improve the performance or characteristics of a device or system. Through the intentional use of feedback, the gain of amplifiers can be made much greater, or more stable, or less sensitive to change, or less dependent on frequency (Chapter 18). On the other hand, amplifier designers must take steps to prevent unintentional feedback from degrading the performance of their products. In this chapter we study the basic concept of feedback, learn how to reduce complicated systems to the basic form, investigate the general benefits of *positive* and *negative* feedback, and then round out our study of circuits by developing the concept of a transfer function.

Amplifier with Feedback

An electronic amplifier is represented by the circuit model of Fig. 9.2a. The transistors and circuit elements have been replaced by equivalent components where R_i and R_o are the input and output resistances, C_o is an output capacitance, and A is a voltage amplification factor. Such an amplifier with feedback can be represented by the model of Fig. 9.2b. The voltage gain of the amplifier alone, called the *forward gain,* is defined by



(a) General amplifier model (b) Idealized model with feedback

Figure 9.2 An electronic amplifier with voltage feedback.

$$\mathbf{G} = \frac{\mathbf{V}_2}{\mathbf{V}_1} \tag{9-1}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are phasors and $\mathbf{G}$ is a complex function of frequency.

The voltage divider consisting of R_1 and R_2 provides a means of tapping off a portion $\mathbf{V}_F$ of the output voltage $\mathbf{V}_2$ and feeding it back into the input. The *feedback factor* is

$$\mathbf{H} = \frac{\mathbf{V}_F}{\mathbf{V}_2} \tag{9-2}$$

In general, $\mathbf{H}$ is a complex quantity although here $\mathbf{H} = R_1/(R_1 + R_2)$ is a real number.

Now the amplifier input $\mathbf{V}_1$ is the sum of the signal $\mathbf{V}_S$ and the feedback voltage $\mathbf{V}_F$ or

$$\mathbf{V}_1 = \mathbf{V}_S + \mathbf{V}_F \tag{9-3}$$

The output voltage is now

$$\mathbf{V}_2 = \mathbf{G}\mathbf{V}_1 = \mathbf{G}(\mathbf{V}_S + \mathbf{V}_F) = \mathbf{G}(\mathbf{V}_S + \mathbf{H}\mathbf{V}_2) = \mathbf{G}\mathbf{V}_S + \mathbf{G}\mathbf{H}\mathbf{V}_2$$

Solving,

$$\mathbf{G}_F = \frac{\mathbf{V}_2}{\mathbf{V}_S} = \frac{\mathbf{G}}{1 - \mathbf{G}\mathbf{H}} \tag{9-4}$$

where $\mathbf{G}_F$ is the gain with feedback.

Equation 9-4 is the basic equation for system gain with feedback, and by studying this equation some of the possibilities of such a system can be foreseen. If the complex quantity **GH** is positive, as it approaches unity the gain increases without limit. If **GH** is just unity, the gain is infinite and there is the possibility of output with no input; the amplifier has become a self-excited *oscillator*. If **GH** is negative, the gain is reduced, but accompanying benefits may offset the loss in gain. For example, if **GH** is very large with respect to unity, $\mathbf{G}_F \cong \mathbf{G}/\mathbf{GH} = 1/\mathbf{H}$, or the gain is independent of the gain of the amplifier itself. (In what way might this be beneficial?)

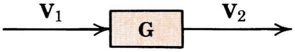
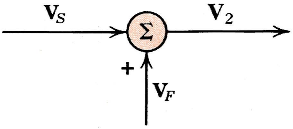
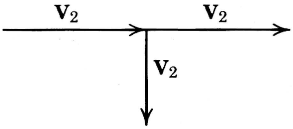
Block Diagrams

Before investigating the specific behavior of feedback systems, we need a simplified notation that will permit us to disregard the details of circuits and devices and focus our attention on system aspects. Also, our approach should emphasize the paths whereby signals are transmitted.

Block Diagram Representation

Let us represent linear two-port devices by blocks labeled to indicate the functions performed and with single-line inputs and outputs labeled to indicate the signals at those points. The amplifier whose small-signal circuit model is within the colored area in Fig. 9.2a is represented symbolically by a rectangle labeled **G** (Table 9-1). Such a symbol can represent any two-port device for which the output signal is a known function of the input signal. In general, the ratio of output signal to input signal is called the *transfer function*.

Table 9-1 Block Diagram Symbols and Defining Equations

Transfer function	Summing point	Pickoff point
	 $\mathbf{V}_S$ Σ $\mathbf{V}_2$ + $\mathbf{V}_F$	
$\mathbf{V}_2 = \mathbf{G}\mathbf{V}_1$	$\mathbf{V}_2 = \mathbf{V}_S + \mathbf{V}_F$	$\mathbf{V}_2 = \mathbf{V}_2 = \mathbf{V}_2$

To be represented by such a block, a device must be linear.† Also, it is assumed that each device has infinite input impedance and zero output impedance, so that devices may be interconnected without loading effects. In the system of Fig. 9.2b, $R_1 + R_2$ must be large with respect to R_L, so that the amplifier gain is unaffected by the feedback connection. If this amplifier is to provide the input to another device that draws an appreciable current, a transistor amplifier, for example, transfer function **G** must be recalculated, taking into account the finite input impedance of the next device. When these conditions are met, the blocks are said to be *unilateral;* linear unilateral blocks may be interconnected freely.

---

† A device describable by a *linear* differential equation in which the variable and its derivatives appear only to the first power.

Two different types of connection are possible (Table 9-1). A *summing point* has two or more inputs and a single output. The output is the sum or difference of the inputs, depending on the signs; a plus sign is assumed if no sign is indicated. In contrast, from a *pickoff point* the signal can be transmitted undiminished in several directions. Physically this is illustrated by a voltage that can serve as input to several devices with high input impedances, or by a tachometer that senses the speed transmitted by a shaft without diminishing the speed.

Using block diagram notation, the amplifier of Fig. 9.2b is represented as in Fig. 9.3. In this case, all variables are voltages but, in general, different variables may appear on the diagram. (All signals entering a summing point must have the same units.) The amplifier block represents Eq. 9-1 and the feedback block represents Eq. 9-2. In this case, **G** is a voltage gain, but in general it is a transfer function. The summing point fulfills Eq. 9-3, and Eq. 9-4 gives the gain of the system. This basic system appears so frequently that Eq. 9-4 should be memorized.

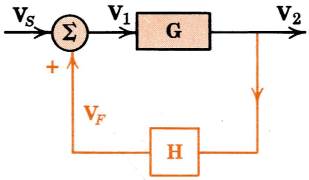


Figure 9.3 Block diagram of the basic feedback system.

Block Diagram Algebra

Analysis of the behavior of complicated systems can be simplified by following a few rules for manipulating block diagram elements. These rules are based on the definitions of block diagram symbols and could be derived each time, but it is more convenient to work from the rules.

1. *Any closed-loop system can be replaced by an equivalent open-loop system.*
2. *The gain of cascaded blocks is the product of the individual gains.*
3. *The order of summing does not affect the sum.*
4. *Shifting a summing point beyond a block of gain G_A requires the insertion of G_A in the variable added.*
5. *Shifting a pickoff point beyond a block of gain G_A requires the insertion of $1/G_A$ in the variable picked off.*

These rules and the corollaries of Rules 4 and 5 are illustrated in Example 1.

Example 1

Determine the overall gain of the system shown in Fig. 9.4a.

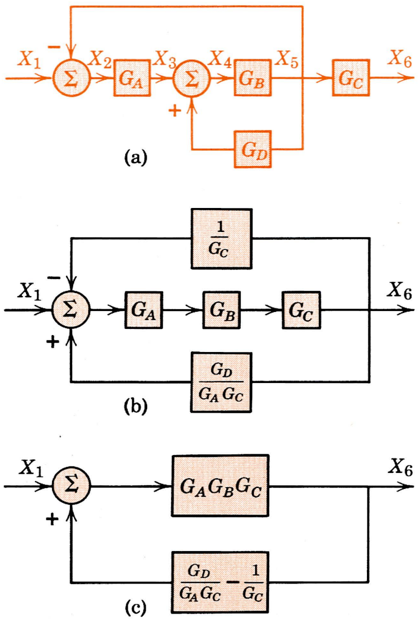


Figure 9.4 Block diagram algebra.

It would be possible to determine the gain by writing the equation for each element in turn and solving simultaneously. A preferable procedure is to simplify the system by block diagram algebra.

Applying rule 5, the pickoff point is moved beyond G_C and $1/G_C$ is inserted in the upper and lower loops.

Applying the corollary of rule 4 permits shifting the summing point for the lower loop before G_A and $1/G_A$ is inserted in that loop (Fig. 9-4b).

Applying rule 3, the two feedback loops are combined.

Applying rule 2, the three cascaded blocks are combined.

The result is in the basic form (Fig. 9-4c).

Applying rule 1, the overall gain can be written by inspection as

$$G_{61} = \frac{X_6}{X_1} = \frac{G_A G_B G_C}{1 - (G_A G_B)(G_D/G_A - 1)}$$

Feedback Circuits

By using the rules of block diagram algebra, all linear feedback networks can be reduced to the basic system of Fig. 9-3 for which the gain is $\mathbf{G}_F = \mathbf{G}/(1 - \mathbf{GH})$. The behavior of the system can be predicted by examining the quantity $1 - \mathbf{GH}$. In a complicated network, the denominator will be a polynomial, as in Example 1, and its poles and zeros will correspond to critical points in the behavior pattern.

Positive Feedback

For **GH** positive but less than unity, the denominator is less than unity and the system gain is greater than the forward gain **G**. This condition of *positive feedback* was used in *regenerative receivers* to provide high gain in the days when electronic amplifiers were poor and signals were weak. Since operation with positive feedback becomes less stable as *GH* approaches unity, regenerative receivers required constant adjustment.

If the value of **GH** approaches $1 + j0$, the denominator approaches zero and the overall gain increases without limit. An amplifier of infinite gain can provide an output with no input; such a device is called an *oscillator* or signal generator and is an important part of many electronic systems.

Oscillation. In the amplifier of Fig. 9.2a, the output tends to decrease as the frequency increases because of the filtering effect of R_0C_0. As we shall see in Chapter 18, the high-frequency gain can be expressed as

$$\mathbf{A} = \frac{A_O}{1 + j\omega C_0 R_0} = \frac{A_O}{1 + j\omega/\omega_2}$$

where A_O is a constant representing the gain at moderate frequencies for which $\omega/\omega_2 \ll 1$, and ω_2 is the *upper cutoff frequency* at which **A** drops to 70% of A_O. At a somewhat higher frequency where $\omega = \sqrt{3}\,\omega_2$,

$$\mathbf{A}_H = \frac{A_O}{1 + j\sqrt{3}} = \frac{A_O}{2}\underline{/-\tan^{-1}\sqrt{3}} = \frac{A_O}{2}\underline{/-60^\circ}$$

For a three-stage amplifier consisting of three similar sections, at this frequency the overall gain is

$$\mathbf{G} = \mathbf{A}_H^3 = \left(\frac{A_O}{2}\right)^3 \underline{/3(-60^\circ)} = \left(\frac{A_O}{2}\right)^3 \underline{/-180^\circ} = -\left(\frac{A_O}{2}\right)^3 + j0$$

Assuming feedback is present and $\mathbf{H} = 0.008$,

$$\mathbf{GH} = -0.008\left(\frac{A_O}{2}\right)^3 + j0 = -\left(\frac{A_O}{10}\right)^3 + j0$$

For oscillation, $\mathbf{G}_F$ must be infinite or the denominator must be zero, or

$$\mathbf{GH} = -\left(\frac{A_O}{10}\right)^3 + j0 = +1 + j0 \quad \text{and} \quad A_O = -10$$

These results are interpreted as follows: In a three-stage amplifier with a per-stage gain of -10, if 0.8% of the output voltage is fed back to the input, oscillation is possible at a frequency for which the per-stage phase shift is 60°. In practice, there is a range of frequencies that will be amplified, but one of these will be maximized and the level of oscillation will build up until A_O is just equal to -10. In this case **G** is the frequency-dependent element; in other cases **H** may determine the frequency at which $\mathbf{GH} = 1 + j0$. (See Chapter 18.)

The preceding illustration indicates that a circuit designed for amplification may produce oscillation if some small part of the output is coupled or fed back to the input with the proper phase relation. Opportunities for such coupling exist in the form of stray electric and magnetic fields or through mutual impedances of common power supplies. Amplifier designers use shielding and filtering to minimize coupling and reduce the possibility of oscillation.

Negative Feedback

Negative feedback, where the amplitude and phase of the signal fed back are such that the overall gain of the amplifier is less than the forward gain, may be used to advantage in amplifier design. Since additional voltage or current gain can be obtained easily with electronic devices, the loss in gain is not important if significant advantages are obtained. Among the possible improvements in amplifier performance are:

1. Gain can be made practically independent of device parameters; gain can be made insensitive to temperature changes, aging of components, and variations from typical values.
2. Gain can be made practically independent of reactive elements and, therefore, insensitive to frequency.
3. Gain can be made selective to discriminate against noise, distortion, or system disturbances.
4. The input and output impedance of an amplifier can be improved greatly.
5. The time constant of a circuit component can be modified to obtain faster response.

Stabilizing Gain. For precise instrumentation or long-term operation, amplifier performance must be predictable and stable. However, amplifier components such as transistors and vacuum tubes vary widely in their critical parameters. By employing negative feedback and sacrificing gain (which can be made up readily) stability can be improved.

Example 2

An amplifier with an initial gain $G_1 = 1000$ undergoes a 10% reduction in gain when a faulty transistor is replaced. Predict the stabilizing effect of a feedback loop where $H = -0.009$.

Using the basic equation for gain with feedback, the initial gain is

$$G_{F1} = \frac{G_1}{1 - G_1H} = \frac{1000}{1 - 1000(-0.009)} = 100$$

If gain is reduced by 10% to $G_2 = 900$,

$$G_{F2} = \frac{G_2}{1 - G_2H} = \frac{900}{1 - 900(-0.009)} = 99$$

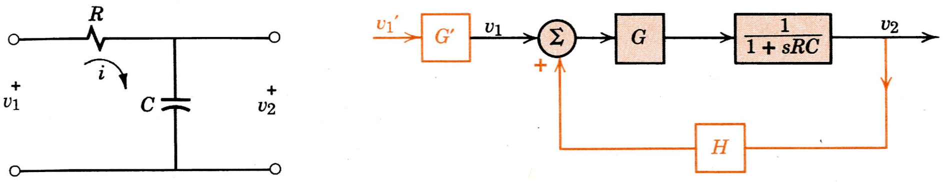
With feedback the change in gain is 1%, or only $\frac{1}{10}$ as great as it is in the amplifier alone.

It can be shown (see Exercise 10) that, for small changes in G, the fractional change in overall gain with feedback is equal to the change in G divided by $1 - GH$. By definition, the *sensitivity* of gain to change is

$$\frac{dG_F/G_F}{dG/G} = \frac{1}{1 - GH} \tag{9-5}$$

In words, the sensitivity of the gain to change is reduced by the factor $1 - GH$.

Reducing Time Constants. The speed of response of control elements is usually a limiting factor in the performance of control systems. An actuating motor with inertia takes a finite time to come up to a given speed, and it takes a finite time for the level in a chemical tank to drop a given amount. By using feedback principles, the speed of response can be greatly increased or, in other words, the time constant of the system can be reduced. The control system applications of this concept are treated in Chapter 26. Here we are interested in the application to electrical circuits.



(a) *RC* two-port

(b) Two-port with feedback

Figure 9.5 Reduction of a time constant by means of feedback.

To illustrate the powerful effect of feedback on the time constant of an electrical circuit, consider the *RC* two-port of Fig. 9.5a. For exponential functions, the transfer function can be derived as

$$\frac{v_2}{v_1} = \frac{(1/sC)i}{(R + 1/sC)i} = \frac{1}{1 + sRC} \tag{9-6}$$

The time constant RC is a measure of the time for v_2 to approach v_1 if v_1 is a suddenly applied dc voltage, for example. (Or it is a measure of the time required for a motor to reach operating speed or a liquid level to drop to a new value.)

If amplification (G) and feedback (H) are provided, the system is described by the block diagram of Fig. 9.5b. For this system,

$$\frac{v_2}{v_1} = \frac{\dfrac{G}{1 + sRC}}{1 - \dfrac{GH}{1 + sRC}} = \frac{G}{1 + sRC - GH} = \frac{G}{1 - GH + sRC} \tag{9-7}$$

If an amplifier is added ahead of v_1 so that

$$G' = \frac{v_1}{v_1'} = \frac{1 - GH}{G}$$

the new transfer function is

$$\frac{v_2}{v_1'} = \frac{v_2}{v_1} \cdot \frac{v_1}{v_1'} = \frac{G}{(1 - GH) + sRC} \cdot \frac{1 - GH}{G} = \frac{1}{1 + s\dfrac{RC}{1 - GH}} \tag{9-8}$$

and the time constant has been reduced by the factor $(1 - GH)$. As in the preceding illustration, a desirable result has been achieved at the expense of amplification.

Transfer Functions

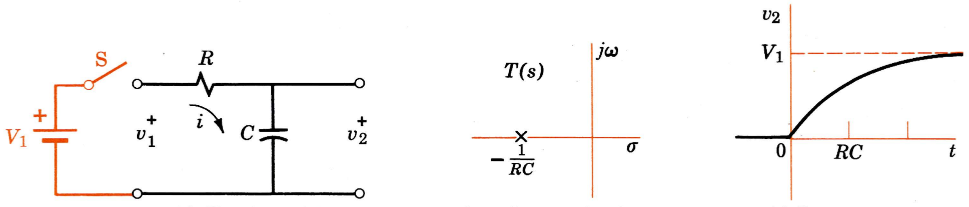
In the previous discussion we defined the transfer function of a two-port as the ratio of output signal to input signal. For exponential signals of the form Ve^{st}, we know that the complete response current in a one-port is determined by the admittance $Y(s)$. Let us now extend that concept and show that the response of a two-port is determined by the transfer function $T(s)$.

The Transfer Function Concept

Applying the voltage-divider principle to the circuit of Fig. 9.6, the ratio of the output voltage to the input voltage is

$$T(s) = \frac{V_2(s)}{V_1(s)} = \frac{1/sC}{R + 1/sC} = \frac{1/RC}{s + 1/RC} \tag{9-9}$$

where $T(s)$ is the voltage transfer function. To determine the output voltage for a step input voltage using our customary procedure (p. 150), we note that for a dc step, $s = 0$, and $T(s) = T(0) = 1$. Therefore, the forced response would be just $v_{2f} = V_1$. Since a pole of the transfer function appears at $s = -1/RC$, we



(a) Circuit model (b) Pole-zero diagram (c) Step response

Figure 9.6 Step response of an RC circuit.

expect the natural response to contain a term $A\,e^{-t/RC}$. The complete response would be

$$v_2 = v_{2f} + v_{2n} = V_1 + A\,e^{-t/RC} \tag{9-10}$$

Assuming the capacitor is initially uncharged, at $t = 0^+$, $v_C = 0$,

$$\therefore v_2 = 0 = V_1 + A\,e^0 \quad \text{and} \quad A = -V_1$$

As shown in Fig. 9.6c, the complete response would be

$$v_2 = V_1 - V_1\,e^{-t/RC} \tag{9-11}$$

To check this result, we note that the complete response current for the one-port is $i = (V_1/R)\,e^{-t/RC}$ since $Y(0) = 0$ and $i_f = 0$. In this circuit, $v_2 = v_C = v_1 - iR$, hence

$$v_2 = V_1 - V_1\,e^{-t/RC}$$

which is identical with the expression obtained by using the transfer function concept.

We conclude from this simple illustration that *the transfer function is derived and interpreted just as an admittance or impedance function.* Note that an immittance is the ratio of two quantities measured at the same port, whereas a transfer function is the ratio of two quantities measured at different ports. The two quantities may be entirely different in physical nature.

For the special case of sinusoidal excitation, the transfer function (Eq. 9-9) becomes

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \mathbf{T}\,(j\omega) = \frac{1}{1 + j\omega CR} = \frac{1}{\sqrt{1 + (\omega CR)^2}} \underline{/-\tan^{-1}\omega CR} \tag{9-12}$$

where $\mathbf{V}_2$ and $\mathbf{V}_1$ are the output and input phasors. It was pointed out in Chapter 3 that the phasor $\mathbf{V}_1 = V_1\,e^{j\theta_1}$ is a *transform* of the sinusoidal function of time $v_1 = V_{1m}\cos(\omega t + \theta_1)$. This transform represents the sinusoidal function. In this chapter let us employ transforms† of exponential functions as well so that $V_1(s)$ might represent a function $v_1 = A_1\,e^{s_1 t} + A_2\,e^{s_2 t} + \cdots$ defined for $t > 0$. Using this notation, we rewrite Eq. 9-9 as

$$\frac{V_2(s)}{V_1(s)} = T(s) = \frac{1/RC}{s + 1/RC} \tag{9-13}$$

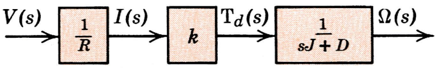
---

† The Laplace transformation is a mathematical operation that can be used to transform functions of time into functions of the complex variable s. Here the transfer function $T(s)$ is the ratio of the Laplace transforms of voltages $v_1(t)$ and $v_2(t)$ where v_1 and v_2 are defined for $t > 0$ and are equal to zero for $t \leq 0$. Powerful methods are available for determining the complete response of a system directly from the transfer function using Laplace transformation. See Gardner and Barnes, *Transients in Linear Systems,* John Wiley & Sons, New York 1942.

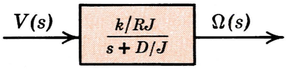
The value of the transfer function concept and the transform notation in control system analysis is illustrated in Example 3.

Example 3

The torque developed in a certain electric motor is directly proportional to magnetic field current i. Part of the developed torque τ_d is used in overcoming friction and the remainder is available for acceleration. Determine the response of the motor to a step input of field voltage V.



(a) Block diagram



(b) Transfer function

Figure 9.7 Transfer function of a motor.

The governing equations (see Table 5-1 p. 135) and the corresponding transfer functions are:

$$i = \frac{v}{R} \qquad \frac{I(s)}{V(s)} = \frac{1}{R}$$

$$\tau_d = ki \qquad \frac{T_d(s)}{I(s)} = k$$

$$\tau_d = J\frac{d\omega}{dt} + D\omega \qquad \frac{\Omega(s)}{T_d(s)} = \frac{1}{sJ + D}$$

The block diagram is as shown in Fig. 9.7a.

The overall transfer function is

$$\frac{\Omega(s)}{V(s)} = \frac{I(s)}{V(s)} \times \frac{T_d(s)}{I(s)} \times \frac{\Omega(s)}{T_d(s)} = \frac{k/R}{sJ + D} = \frac{k/RJ}{s + D/J}$$

For $s = 0$, $\Omega(0) = k/RD$ and, by analogy with Eq. 9-11,

$$\omega(t) = \frac{k}{RD} - \frac{k}{RD}e^{-(D/J)t}$$

The motor speed ω approaches as a limit the speed governed by rotational friction D alone.

Advantages in Using Transfer Functions

In systems analysis, one possible approach is to write a set of differential equations and solve them simultaneously. A more convenient approach is based on the transfer function concept. The following observations are based on Figs. 9-1 and 9-7 and the accompanying discussion:

Systems usually consist of combinations of cascaded elements, and block diagram representation is convenient.

If each element is characterized by its transfer function, an overall transfer function can be determined by using the rules of block diagram algebra.

The individual transfer functions consist of relatively simple algebraic factors and the overall function is just a combination of these.

Analogous transfer functions are used in characterizing electrical, mechanical, hydraulic, and pneumatic elements and the interpretation of one is applicable to all.

The transfer function approach permits the determination of the transient response or the steady-state sinusoidal response.

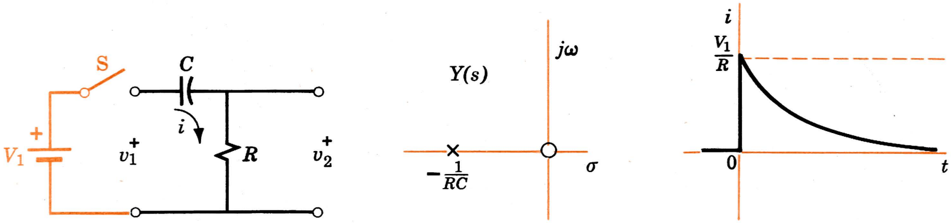
Looking at this list, we can understand why the transfer function approach is used universally.

Dynamic Response

To be useful a control system must be able to cope with changing conditions such as a sudden change in load or a disturbance at some point in the system. In evaluating a system, one approach is to determine the dynamic behavior described by the complete response to a step function. An alternative approach is to consider the frequency response for steady-state sinusoidal inputs. In either approach we rely on the transfer function.

Since transfer functions are derived from linear ordinary differential equations, devices characterized by transfer functions must be represented by models made up of lumped, linear elements. The circuit models we have used to represent electrical, and mechanical devices satisfy this requirement, so transfer functions can be determined directly as in Example 3. Once the transfer function is obtained, the complete step response can be determined following the procedure outlined in Chapter 6.

First-Order Systems. The order of a system is defined by the highest power of s in the denominator of the transfer function. For the RC circuit of Fig. 9.8, the transfer function is (using the voltage-divider principle)



(a) Circuit model (b) Pole-zero diagram (c) Step response

Figure 9.8 Dynamic response of a first-order system.

$$T(s) = \frac{V_2(s)}{V_1(s)} = \frac{R}{R + 1/sC} = \frac{s}{s + 1/RC} \tag{9-14}$$

Since s appears in the denominator to the first power only, this is a first-order system. For $s = 0$, $T(s) = T(0) = 0$ and the forced response is zero. A pole of the transfer function appears at $s = -1/RC$, so the natural response contains a term $A\,e^{-t/RC}$.

Assuming the capacitor is initially uncharged, at $t = 0^+$, $v_C = 0$, and

$$v_2 = V_1 = A_1\,e^0$$

As shown in Fig. 9.8c, the voltage response is iR or

$$v_2 = V_1 e^{-t/RC} \tag{9-15}$$

First-order transfer functions of the form of Eq. 9-14 and Eq. 9-9 occur frequently in control systems, and it is worthwhile to generalize from the results obtained in these two specific cases. For a first-order system having a transfer function of the form

$$T(s) = \frac{ks}{s + \alpha} \tag{9-16}$$

the step response $y(t)$ is always of the form

$$y(t) = Y e^{-\alpha t} \tag{9-17}$$

For a system having a transfer function of the form

$$T(s) = \frac{k}{s + \alpha} \tag{9-18}$$

the step response $y(t)$ is always of the form

$$y(t) = Y(1 - e^{-\alpha t}) \tag{9-19}$$

The presence of a zero at $s = 0$ in the first transfer function (Eq. 9-16) completely changes the step response.

Second-Order Systems. In Example 3 it was assumed that field current i was directly proportional to v and limited only by resistance R. Actually, the winding used to produce the magnetic field necessary for motor operation has a high inductance. How does the motor respond if this inductance is taken into account? The answer to this question is found in the form of the new transfer function, and its derivation illustrates the convenience of our approach.

The new relation for the field circuit is $v = (R + sL)i$ and the transfer function relating field current and voltage (actually an admittance) is

$$T_1(s) = \frac{I(s)}{V(s)} = \frac{1}{sL + R} \tag{9-20}$$

For the motor as a transducer, the transfer function relating developed torque and field current is

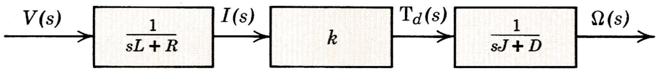
$$T_2(s) = \frac{\mathrm{T}_d(s)}{I(s)} = k \tag{9-21}$$

For the mechanical side of the motor, the transfer function relating shaft velocity and developed torque is

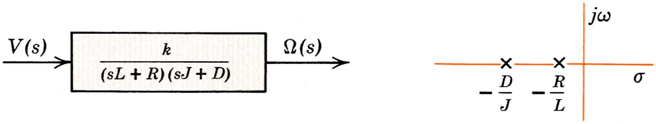
$$T_3(s) = \frac{\Omega(s)}{\mathrm{T}_d(s)} = \frac{1}{sJ + D} \tag{9-22}$$

These three transfer functions are shown in Fig. 9.9a.

Applying the rules of block diagram algebra, we obtain the overall function



(a) Block diagram



(b) Overall transfer function (c) Pole-zero diagram

Figure 9.9 A field-excited shunt motor, a second-order system.

$$T(s) = \frac{I(s)}{V(s)} \times \frac{T_d(s)}{I(s)} \times \frac{\Omega(s)}{T_d(s)} = \frac{\Omega(s)}{V(s)} = \frac{k}{(sL + R)(sJ + D)} \tag{9-23}$$

What is the character of the step response of this system? There are two poles of the transfer function corresponding to two roots of the denominator (Fig. 9.9c). In other words, since s appears in the denominator to the second power, this is a second-order system. Combining the forced response with the natural response, we anticipate that the total step response is of the form

$$\omega(t) = \frac{k}{RD} V + A_1 e^{-Rt/L} + A_2 e^{-Dt/J} \tag{9-24}$$

Even without evaluating A_1 and A_2, we can see that in this case there are two time constants influencing the response. The fact that field current builds up slowly tends to make the response sluggish. Also, the inertia of the motor and load limits the rate of change of shaft velocity. Since a rapid response is usually desired in a control system, this is not a desirable mode of operation.

In this case, the two poles are real and negative. From our experience with electric circuits, we expect that in another case there might be complex roots and an entirely different form of response. The other cases are considered in detail in Chapter 26.

Frequency Response

Another useful interpretation of the transfer function is in terms of the steady-state response, i.e., amplitude and phase angle as functions of frequency. In some cases, a laboratory measurement of frequency response is the best method of determining the system characteristics. Given a transfer function, the frequency response is obtained directly by substituting $s = j\omega$. For the system described by $T(s) = 1/(s\tau + 1)$ where τ is the *time constant*, the frequency response is

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \mathbf{T} = \frac{1}{1 + j\omega\tau} = \frac{1}{\sqrt{1 + (\omega\tau)^2}} \underline{/\tan^{-1} - \omega\tau} \tag{9-25}$$

The calculation of overall gain in complex systems or multistage amplifiers is simplified by using a logarithmic unit. By definition, the power gain in *bels* (named in honor of Alexander Graham Bell) is the logarithm to the base 10 of the power ratio. A more convenient unit is the *decibel* where

$$\text{Gain in decibels} = \text{dB} = 10 \log \frac{P_2}{P_1} \tag{9-26}$$

Since for a given resistance, power is proportional to the square of the voltage,

$$\text{Gain in decibels} = 10 \log \frac{V_2^{\,2}}{V_1^{\,2}} = 20 \log \frac{V_2}{V_1} \tag{9-27}$$

Example 4

Express in decibels (dB) the relative response of a resonant circuit at the lower half-power frequency.

Since this is a half-power frequency, the relative response in dB is

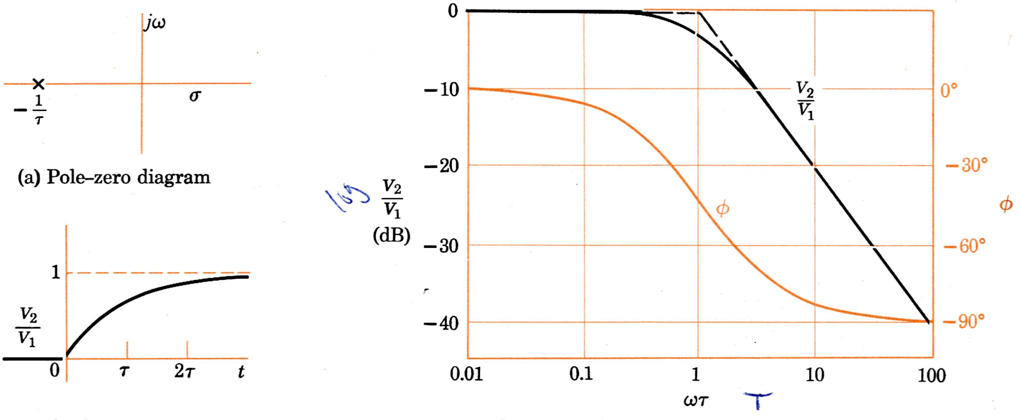
$$10 \log \frac{P_1}{P_0} = 10 \log 0.5 \cong -3 \text{ dB}$$

Alternatively, at f_1 the response is 0.707 of the resonant frequency value. Therefore, the relative gain in dB is

$$20 \log \frac{V_1}{V_0} = 20 \log 0.707 \cong -3 \text{ dB}$$

The response is "down 3 dB" at the cutoff frequencies.

If we follow the procedure of H. W. Bode and plot gain in decibels using the gain at zero frequency as a base, the *Bode chart* of Eq. 9-25 is as shown in Fig. 9.10c. For $\omega\tau \gg 1$, the gain is inversely proportional to $\omega\tau$ and the gain



(a) Pole–zero diagram

(b) Step response

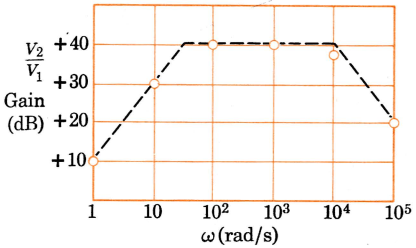
(c) Frequency response

Figure 9.10 Interpretations of the transfer function $T(s) = 1/(s\tau + 1)$.

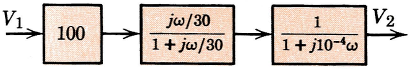
curve on this log-log plot is a straight line with a slope of -20 dB/decade. The gain curve can be approximated by asymptotes that intersect at the *breakpoint* or *corner frequency*. At the breakpoints the phase angle is $-45°$ and the gain is actually down 3 dB. One important virtue of the Bode chart is that the response curve for the *product* of two transfer functions is just the *sum* of the two response curves when gains are plotted in dB.

Example 5

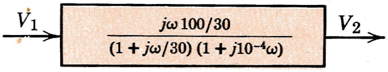
Laboratory data for the response curve of an amplifier are plotted as small circles in Fig. 9-11a. Determine the transfer function of the amplifier.



(a) Frequency response



(b) Block diagram



(c) Transfer function

Figure 9.11 Amplifier performance in terms of the transfer function.

It appears that at high and low frequencies the response drops off at about 20 dB per decade. On the basis of this evidence (and some experience!), the dashed straight-line approximation is drawn. The high-frequency response with a breakpoint ($\omega\tau_2 = 1$) at 10^4 rad/s corresponds to a transfer function of the form

$$\mathbf{T}_2 = \frac{1}{1 + j\omega\tau_2} = \frac{1}{1 + j(\omega/\omega_2)} = \frac{1}{1 + j10^{-4}\omega}$$

The low-frequency response indicates that, for $\omega\tau_1 \ll 1$, the gain is directly proportional to frequency. With a breakpoint ($\omega\tau_1 = 1$) at 30 rad/s, this corresponds to a transfer function of the form

$$\mathbf{T}_1 = \frac{j\omega\tau_1}{1 + j\omega\tau_1} = \frac{j\omega/30}{1 + j\omega/30}$$

For the middle frequencies around $\omega = 10^3$, $T_2 \cong 1$ and $T_1 \cong 1$. By Eq. 9-26 the mid-frequency gain of the amplifier is

$$A = \frac{V_2}{V_1} = \text{antilog}\,\frac{40}{20} = 10^2 = 100$$

The amplifier can be represented by the three ideal elements shown in Fig. 9.11b or by the single transfer function

$$\mathbf{T} = \frac{j\omega(100/30)}{(1 + j\omega/30)(1 + j10^{-4}\omega)} \tag{9-28}$$

SUMMARY

- A system is a combination of diverse but interacting elements integrated to achieve an overall objective.
  Feedback is important in biological, social, and technical systems.

- In a closed-loop system, a portion of the output is fed back and compared to or combined with the input. The gain of the basic closed-loop system is

$$G_F = \frac{G}{1 - GH}$$

- Complicated systems that can be analyzed into linear, unilateral two-port elements are conveniently represented by simple diagrams consisting of blocks, summing points, and pickoff points.
  Block diagrams can be manipulated according to a set of simple rules.

- The quantity $1 - GH$ is the key to the behavior of closed-loop systems.
  For $|1 - GH| < 1$, positive feedback occurs and the system gain is greater than the forward gain.
  For $|1 - GH| = 0$, the system gain increases without limit and oscillation is possible.
  For $|1 - GH| > 1$, negative feedback occurs and the system gain is less than the forward gain.

- An oscillator is a self-excited amplifier in which a portion of the output signal is fed back in the correct phase and amplitude to make $\mathbf{GH} = 1 + j0$.
  By making G or H frequency dependent, oscillation at a single frequency can be obtained.

- The benefits of negative feedback outweigh the loss in gain:
  Gain can be made insensitive to device parameters.
  Gain can be made insensitive to frequency.
  Time constants of components can be modified.
  Noise and distortion can be discriminated against.
  Input and output impedances can be improved.

- The transfer function is the ratio of two exponential functions of time; it is derived and interpreted just as an immittance function.
  The transform $X(s)$ represents the exponential function $x(t)$ for $t > 0$.
  Transfer functions can be determined directly from circuit models.
  Transfer functions follow the rules of block diagram algebra.

- Frequency response is obtained directly from the transfer function.
  Step response is the sum of forced response (for $s = 0$) and natural response determined by the poles of the denominator of $T(s)$.

- The system transfer function may be written in terms of pole and zero locations or in terms of steady-state characteristics ($s = j\omega$).
  On a Bode chart (dB versus log frequency), the response curve for the product of two transfer functions is the sum of the individual response curves.

REVIEW QUESTIONS

1. Cite an example of feedback in a biological system and describe its operation.
2. Cite an example of feedback in a social system and describe its operation.
3. Cite an example of feedback in each of the following branches of engineering: aeronautical, chemical, civil, industrial, and mechanical.
4. If the speaker of a public address system is moved farther away from the microphone, what is the effect on the possibility of a "squeal," the frequency of the squeal if it does develop, and the amplitude of the squeal?
5. Differentiate between the two general categories of feedback applications in electrical systems described in the introduction to this chapter.
6. From memory, draw and label a basic feedback system and write the expression for overall gain.
7. What is the physical meaning for $GH = 1$? What is the ac *power* gain around the closed loop for this situation?
8. List three conditions that must be met if a system is to be represented by a block diagram.
9. Write out the corollaries to block diagram algebra rules IV and V.
10. Draw a complicated system with two pick-off points and two summing points. Using the rules of block diagram algebra, reduce the system to an open-loop equivalent.
11. Define positive feedback and negative feedback.
12. Explain how a multistage audio amplifier may become an oscillator at a certain frequency.
13. List five specific benefits that may result from negative feedback and indicate where they would be important.
14. List three reasons why the gain of a transistor amplifier without feedback might change in normal operation.
15. Explain on a physical basis the operation of feedback in reducing the time constant in Fig. 9.5. Consider voltages at various times at various points.
16. What information on system performance is provided by step response?
17. Define "transform," "transfer function," and "step response."
18. Outline the procedure for determining step response from the transfer function.
19. What information is obtained from the poles of $T(s)$? The zeros?
20. What form of step response is indicated by complex poles of $T(s)$?

EXERCISES

1. A driver on a freeway wishes to keep his left fender two feet from the white line. Draw a labeled block diagram for the closed-loop system he uses.
2. Given a system described by the following equations:

$$V_2 = V_1 + V_4 - AV_3$$
$$V_3 = BV_2 \qquad V_4 = CV_3$$

where V_1 is the input and V_4 is the output.
(a) Draw a labeled block diagram representing this system.
(b) Determine the gain V_4/V_1.

3. Given a system described by the following equations:

$$I_2 = I_1 + V_3/R_3 \qquad I_3 = V_3/B$$
$$V_2 = R_2 I_2 \qquad V_4 = AV_3 \qquad V_3 = V_2 - V_4$$

where I_1 is the input and I_3 is the output.
(a) Draw a labeled block diagram representing this system.
(b) Determine the gain I_3/I_1.

4. Given a system described by the following equations:

$$V_2 = V_1 - I_1R_1 \qquad I_1 = I_2 + I_3$$
$$V_2 = I_2R_2 = I_3R_3$$

where V_1 is the input and I_3 is the output.
(a) Draw a block diagram representing this system.
(b) Draw the corresponding electrical circuit.
(c) Calculate I_3/V_1 from the block diagram (by first reducing it to the basic system of Fig. 9.3), and then from the circuit.

5. Reduce the block diagram of Fig. 9.12 to the basic feedback system of Fig. 9.3 and indicate the values of G and H. Determine the gain of the system.

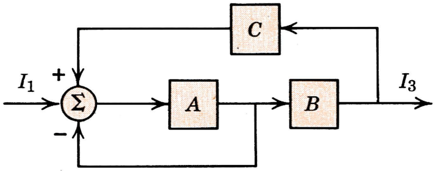


Figure 9.12

6. Repeat Exercise 5 for the diagram of Fig. 9.13.

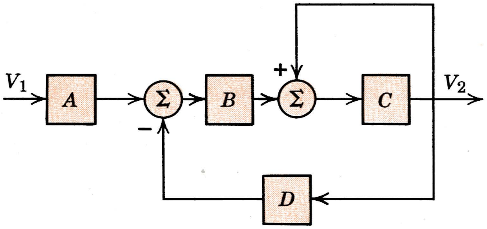


Figure 9.13

7. A feedback system is represented by Fig. 9.3 with $\mathbf{H} = 0.005\underline{/0^\circ}$. Define the conditions necessary for oscillation.

8. A feedback system is represented by Fig. 9.12 with $\mathbf{A} = 50\underline{/90^\circ}$ and $\mathbf{B} = 20\underline{/90^\circ}$. Determine the value of $\mathbf{C}$ necessary for oscillation.

9. For frequencies below $\omega = 2000$ rad/s, the gain of an amplifier is described by $\mathbf{G} = 0.1\omega\underline{/180^\circ}$. For oscillation at $\omega = 1500$ rad/s, what value of $\mathbf{H}$ is required?

10. Consider the basic feedback system of Fig. 9.3. Assuming small changes in G, derive a general expression for the fractional change in overall gain with feedback (dG_F/G_F) in terms of the change in forward gain (dG/G) and the values of G and H.

11. Given the circuit in Fig. 9.14,
- (a) Define and derive the transfer function.
- (b) Sketch the pole-zero diagram of $T(s)$.
- (c) Sketch the step response.

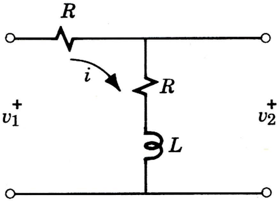


Figure 9.14

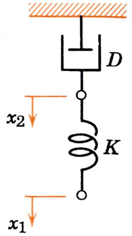


Figure 9.15

12. Repeat Exercise 11 for the system of Fig. 9.15.

13. Repeat Exercise 11 for the system of Fig. 9.16.

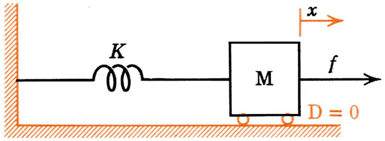


Figure 9.16

14. Given $T(s) = (s + 2a)/(s + a)$,
- (a) Sketch the pole-zero diagram.
- (b) Determine the step response.

15. Derive the transfer function $T(j\omega)$ for the system of Fig. 9.16. For $M = 0.1$ kg and $K = 0.4$ m/N, sketch the amplitude A and the phase angle ϕ as functions of frequency ω.

16. Determine the response to a step of amplitude 10 (no initial energy storage) given:
- (a) $T(s) = 2s/(s + 5)$
- (b) $T(s) = 5/(s + 2)$

17. Determine the transfer function given the unit step response:
- (a) $x(t) = 5 - 5\,e^{-2t}$
- (b) $i(t) = 6\,e^{-t/50}$

18.
- (a) Explain why the natural response of a system is defined by the poles of the transfer function rather than the zeros.
- (b) Given $T(s) = (s + a)/(s^2 + sb)$, write an expression for the natural response.

19. Plot on a Bode diagram (gain in decibels referred to $\omega = 0$) for $0.1 < \omega < 1000$ rad/s:
- (a) $G_a = 1/(1 + j\omega)$
- (b) $G_b = 1/(20 + j2\omega)$
- (c) $G_c = 1/(1 + j\omega)(20 + j2\omega)$

20. The frequency response of a transfer function $T(j\omega)$ is given in Fig. 9.17.

(a) Determine the transfer function.

(b) Determine the step response, assuming no initial energy storage.

21. The power loss in 1 km of coaxial telephone cable is 1.5 dB.

(a) What is the power loss in 2 km of telephone cable? Three?

(b) What is the total power loss from New York to Los Angeles?

(c) What signal power input is required at New York to deliver 1 μW at Los Angeles? Would amplifiers be desirable?

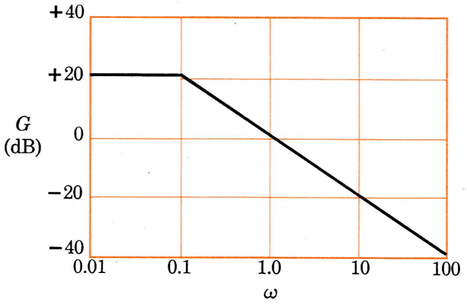


Figure 9.17

PROBLEMS

1. An amplifier with a gain $G_1 = 200$ is subject to a 20% change in gain due to part replacement. Design an amplifier (in block diagram form) that provides the same overall gain but in which similar part replacement will produce a change in gain of only 0.1%. Use amplifier G_1 as part of your design.
2. A cheap amplifier ordered from a catalog is supposed to have a voltage gain of 100, but when delivered proves to have a gain of 70. In fact, when first turned on it has a gain of only 50, which rises to 70 over the course of a minute or two. A second amplifier of the same type shows an initial gain of 90, rising to 110 during warmup.

   Design and diagram an arrangement to obtain a fairly stable gain of 100 by cascading the two amplifiers. Specify the feedback factor. Predict the percentage change in gain of your amplifier system during warmup.
3. The low-frequency response of an amplifier is defined by $\mathbf{A} = A_O/(1 - j\omega_1/\omega)$ where $\omega_1 = 200$ rad/s and $A_O = 200$.
   (a) Plot a couple of points and sketch a graph of A/A_O versus ω with ω on a logarithmic scale.
   (b) Assume that a feedback loop with $H = -0.04$ is provided and repeat part (a). Label this graph "With feedback."
4. For the mechanical system in Fig. 9.18,
   (a) Derive the transfer function relating position x_2 to force f.

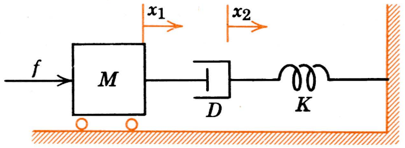


Figure 9.18

   (b) For $M = 1$ N, $D = 1$ N-s/m, and $K = 0.5$ m/N, sketch the pole-zero diagram and the step response.
5. In order to reduce the time constant of the circuit in Fig. 9.19, a fraction k of the voltage across R is to be fed back. The feedback voltage is combined with a new input voltage V_0 and introduced into an amplifier of gain A which supplies V_1.
   (a) Draw a labeled wiring diagram and a block diagram showing the system with feedback.
   (b) Determine the step response of the system and compare the time constants with and without feedback.
   (c) Design the system (specify V_0, k, and A) so that the step response with feedback has the same amplitude but is 10 times as fast.

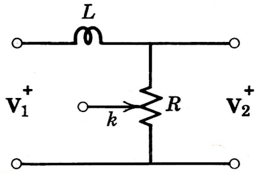


Figure 9.19

PART 2

ELECTRONIC DEVICES

Chapter 10
Electron Tubes

Electron Motion
Cathode-Ray Tubes
Vacuum Diodes
Two-Port Devices

The history of electronics starts with the discovery of cathode rays and continues into tomorrow. It is a fascinating story of individual contributions by mathematicians, physicists, engineers, and inventors and of the painstaking building of a solid technology.

While Hittorf and Crookes were studying cathode rays (1869), Maxwell was developing his mathematical theory of electromagnetic radiation. Soon after Edison observed electronic conduction in a vacuum (1883), Hertz demonstrated (1888) the existence of the radio waves predicted by Maxwell. At the time of J. J. Thomson's measurement of e/m (1897), Marconi was becoming interested in wireless and succeeded in spanning the Atlantic (1901). While Einstein was generalizing from the photoelectric effect, Fleming was inventing the first electron tube (1904), a sensitive diode detector utilizing the Edison effect. DeForest's invention of the triode (1906) made it possible to amplify signals electronically and led to Armstrong's sensitive regenerative detector (1912) and the important oscillator.

Zworykin's invention of the picture tube (1924) and the photomultiplier (1939) opened new areas of electronics. Watson-Watts' idea for radio detection and ranging, or *radar,* developed rapidly under the pressure of World War II. Postwar demands and increased knowledge of semiconductor physics led to the invention of the transistor by Shockley, Bardeen, and Brattain (1947) and the silicon solar cell by Pearson (1954). Kilby's integrated circuit (1958) permitted placing a complete network containing many semiconductor devices on a single monolithic chip. Einstein had pointed out (1917) the possibility of stimulated emission of radiation, and Townes used this concept (1953) to achieve *maser* operation. Townes and Schawlow (1958) described a device that could extend the maser principle into the optical region, and Maiman succeeded in doing so (1960) with his pulsed ruby *laser.*

In this chapter we take a first look at some of the physical principles that underlie the operation of all electronic devices. To design new devices much more study of fundamental principles will be necessary. Our aim here is to gain the understanding required for effective use of existing devices in conventional applications. First, we examine the motion of individual electrons in a vacuum and derive the equations applicable to the cathode-ray tube. Then we consider the effect of many electrons, as in a practical vacuum tube. Finally, we analyze the characteristics of some important electronic devices and see how amplification is achieved.

Electron Motion

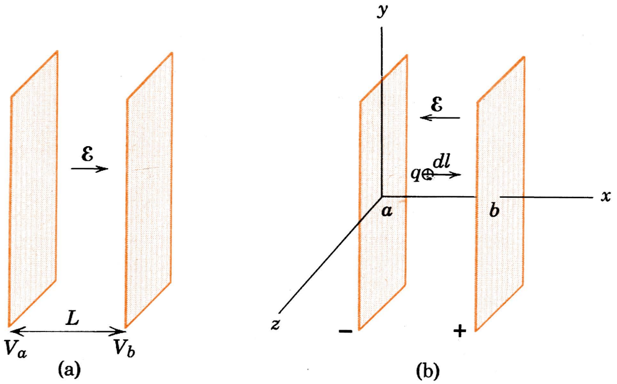
The model of the electron as a negatively charged particle of finite mass but negligible size is satisfactory for many purposes. Based on many careful measurements, the accepted values for charge and mass of the electron are

$$e = 1.602 \times 10^{-19}\,\mathrm{C} \cong 1.6 \times 10^{-19}\,\mathrm{C}$$
$$m = 9.109 \times 10^{-31}\,\mathrm{kg} \cong 9.1 \times 10^{-31}\,\mathrm{kg}$$

In contrast, the hydrogen ion, which carries a positive charge of the same magnitude, has a mass approximately 1836 times as great. If the mass, charge, and initial velocity are known, the motion of individual electrons and ions in electric and magnetic fields can be predicted using Newton's laws of mechanics.

Motion in a Uniform Electric Field

A uniform electric field of strength $\mathcal{E}$ is established between the parallel conducting plates of Fig. 10.1a by applying a potential difference or voltage.



(a) (b)

Figure 10.1 An electric charge in a uniform electric field.

By definition (Eq. 1-5)

$$\mathcal{E} = -\frac{dv}{dl} = -\frac{V_b - V_a}{L} \text{ volts/meter} \tag{10-1}$$

if the spacing is small compared to the dimensions of the plates. Note that if $V_b > V_a$, $V_b - V_a$ is a positive quantity and the electric field is negative (directed to the left in Fig. 10.1b).

By definition (Eq. 1-4), the electric field strength is the force per unit positive charge. Therefore, the force in newtons on a charge q in coulombs is

$$\mathrm{f} = q\mathcal{E} \tag{10-2}$$

Considering the energy dw gained by a charge q moving a distance dl against the force of the electric field (Eq. 1-3), the voltage of point b with respect to point a is

$$V_{ba} = \frac{1}{q}\int_a^b dw = \frac{1}{q}\int_a^b f\,dl = \frac{1}{q}\int_a^b q(-\mathcal{E})\,dl = -\int_a^b \mathcal{E}\,dl \tag{10-3}$$

In words, the voltage between any two points in an electric field is the line integral of the electric field strength. Equation 10-3 is the corollary of Eq. 10-1.

In general, an electron of charge $-e$ in an electric field $\mathcal{E}$ experiences a force

$$f_x = (-e)(\mathcal{E}) = -e\mathcal{E} = ma_x$$

and an acceleration

$$a_x = -\frac{e\mathcal{E}}{m} \tag{10-4}$$

We see that an electron in a uniform electric field moves with a constant acceleration. We expect the resulting motion to be similar to that of a freely falling mass in the earth's gravitational field. Where u is velocity and x is displacement, the equations of motion are:

$$u_x = \int_0^t a_x\,dt = a_x t + U_0 = -\frac{e\mathcal{E}}{m}t + U_0 \tag{10-5}$$

$$x = \int_0^t u_x\,dt = \frac{a_x t^2}{2} + U_0 t + X_0 = -\frac{e\mathcal{E}}{2m}t^2 + U_0 t + X_0 \tag{10-6}$$

(See Example 1 on p. 258.)

Energy Gained by an Accelerated Electron

When an electron is accelerated by an electric field it gains kinetic energy at the expense of potential energy, just as does a freely falling mass. Since voltage is energy per unit charge, the potential energy "lost" by an electron in "falling" from point a to point b is, in joules,

$$\mathrm{PE} = W = q(V_a - V_b) = -e(V_a - V_b) = eV_{ba} \tag{10-8}$$

Example 1

A voltage V_D is applied to an electron deflector consisting of two horizontal plates of length L separated a distance d, as in Fig. 10.2. An electron with initial velocity U_O in the positive x direction is introduced at the origin. Determine the path of the electron and the vertical displacement at the time it leaves the region between the plates.

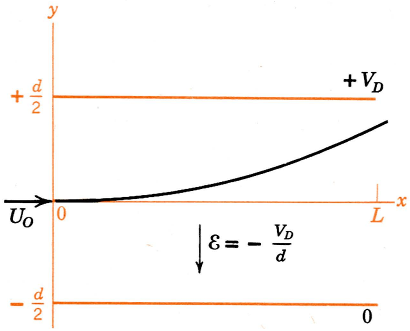


Figure 10.2 Calculation of electron deflection in a uniform electric field.

Assuming no electric field in the x direction and a uniform field of magnitude $\mathcal{E}_y = -V_D/d$ in the negative y direction, the accelerations are:

$$a_x = 0 \quad \text{and} \quad a_y = -\frac{e\mathcal{E}_y}{m} = \frac{eV_D}{md}$$

There is no acceleration in the x direction and the electron moves with constant velocity to the right. There is a constant upward acceleration and the electron gains a vertical component of velocity. The path is determined by:

$$x = U_O t \quad \text{and} \quad y = -\frac{e\mathcal{E}}{2m}t^2 = \frac{eV_D}{2md}t^2$$

Eliminating t,

$$y = \frac{eV_D}{2mdU_O{}^2}x^2 \tag{10-7}$$

or the electron follows a parabolic path.

At the edge of the field, $x = L$ and the vertical displacement is

$$y = \frac{eV_D}{2mdU_O{}^2}L^2$$

If V_D exceeds a certain value, displacement y exceeds $d/2$ and the electron strikes the upper plate.

where V_{ba} is the potential of b with respect to a.

The kinetic energy gained, evidenced by an increase in velocity, is just equal to the potential energy lost, or

$$\text{KE} = \tfrac{1}{2}mu_b{}^2 - \tfrac{1}{2}mu_a{}^2 = \text{PE} = eV_{ba} \tag{10-9}$$

This important equation indicates that the kinetic energy gained by an electron in an electric field is determined only by the voltage difference between the initial and final points; it is independent of the path followed and the electric field configuration. (We assume that the field does not change with *time*.)

Frequently we are interested in the behavior resulting from a change in the energy of a single electron. Expressed in joules, these energies are very small; a more convenient unit is suggested by Eq. 10-8. An *electron volt* is the potential energy lost by 1 electron falling through a potential difference of 1 volt. By Eq. 10-8,

$$1\ \text{eV} = (1.6 \times 10^{-19}\ \text{C})(1\ \text{V}) = 1.6 \times 10^{-19}\ \text{J} \tag{10-10}$$

For example, the energy required to remove an electron from a hydrogen atom is about 13.6 eV. The energy imparted to an electron in a linear accelerator may be as high as 20 BeV (20 billion electron volts).

For the special case of an electron starting from rest ($u_a = 0$) and accelerated through a voltage V, Eq. 10-9 can be solved for $u_b = u$ to yield

$$u = \sqrt{2(e/m)V} = 5.93 \times 10^5 \sqrt{V} \text{ m/s} \tag{10-11}$$

In deriving Eq. 10-11 we assume that mass m is a constant; this is true only if the velocity is small compared to the velocity of light, $c \cong 3 \times 10^8$ m/s.

Example 2

Find the velocity reached by an electron accelerated through a voltage of 3600 V.

Assuming that the resulting velocity u is small compared to the velocity of light c, Eq. 10-11 applies and

$$u = 5.93 \times 10^5 \sqrt{3600} = 3.56 \times 10^7 \text{ m/s}$$

In this case,

$$\frac{u}{c} = \frac{3.56 \times 10^7}{3 \times 10^8} \cong 0.12$$

At the velocity of Example 2 the increase in mass is appreciable, and the actual velocity reached is about 0.5% lower than that predicted. For voltages above 4 or 5 kV, a more precise expression should be used (see Problem 1).

Motion in a Uniform Magnetic Field

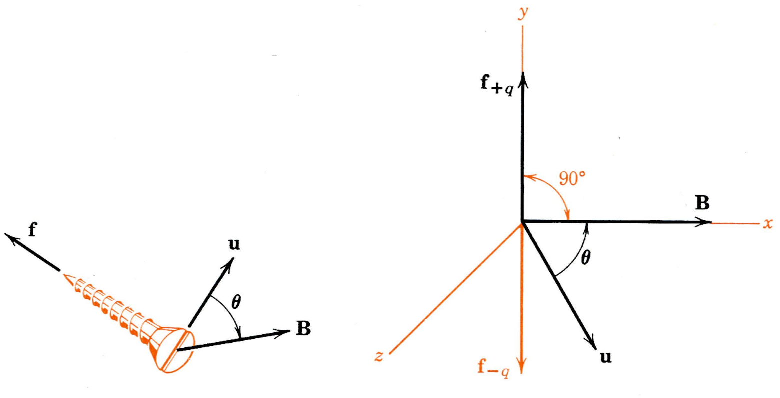
One way of defining the strength of a magnetic field (Eq. 1-6) is in terms of the force exerted on a unit charge moving with unit velocity normal to the field. In general, the force in newtons is

$$\mathbf{f} = q\mathbf{u} \times \mathbf{B} \tag{10-12}$$

where q is charge in coulombs and B is magnetic flux density in teslas (or webers/meter^2). The vector cross product is defined by the right-hand screw rule illustrated in Fig. 10.3. Rotation from the direction of **u** to the direction of **B** advances the screw in the direction of **f.** The magnitude of the force is $quB \sin \theta$ and the direction is always normal to the plane of **u** and **B.**

Equation 10-12 is consistent with three observable facts:

1. A charged particle at rest in a magnetic field experiences no force ($u = 0$).
2. A charged particle moving parallel with the magnetic flux experiences no force ($\theta = 0$).



(a) (b)

Figure 10.3 The right-hand screw rule defining the vector cross product.

3. A charged particle moving with a component of velocity normal to the magnetic flux experiences a force that is normal to u and therefore the magnitude of velocity (or speed) is unchanged.

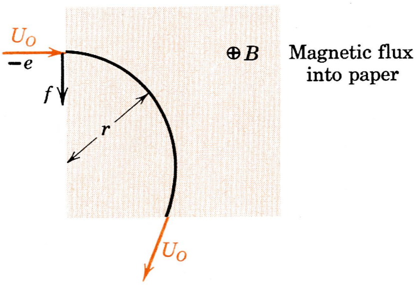


Figure 10.4 Electron motion in a uniform magnetic field.

From the third statement we conclude that no work is done by a magnetic field on a charged particle and its kinetic energy is unchanged.

Figure 10.4 shows an electron entering a finite region of uniform flux density. For an electron ($q = -e$) moving in the plane of the paper $\theta = 90°$), the force is in the direction shown with a magnitude

$$f = eU_O B \qquad (10\text{-}13)$$

Applying the right-hand rule, we see that the initial force is downward and, therefore, the acceleration is downward and the path is deflected as shown. A particle moving with constant speed and constant normal acceleration follows a circular path. The centrifugal force due to circular motion must be just equal to the centripetal force due to the magnetic field or

$$m\frac{U_O{}^2}{r} = eU_O B$$

and the radius of the circular path is

$$r = \frac{mU_O}{eB} \qquad (10\text{-}14)$$

The dependence of radius on the mass of the charged particle is the principle underlying the *mass spectrograph* for studying isotopes.

Motion in Combined $\mathcal{E}$ and B Fields

In the general case, both electric and magnetic fields are present and exert forces on a moving charge. The total force is

$$\mathbf{f} = q(\boldsymbol{\mathcal{E}} + \mathbf{u} \times \mathbf{B}) \tag{10-15}$$

The special cases previously described can be derived from this general relation. If both $\mathcal{E}$ and B are present, the resulting motion depends on the relative orientation of $\mathcal{E}$ and B, and also on the initial velocities. An interesting case is that in which an electron starts from rest in a region where $\mathcal{E}$ and B are mutually perpendicular. The electron is accelerated in the direction of $-\mathcal{E}$, but as soon as it is in motion there is a reaction with B and the path starts to curve. Curving around, the electron is soon traveling in the $+\mathcal{E}$ direction and experiences a decelerating force that brings it to rest. Once the electron is at rest, the cycle starts over; the resulting path is called a *cycloid* and resembles the path of a point on a wheel as it rolls along a line.

If the fields are not uniform, mathematical analysis is usually quite difficult. A case of practical importance is the "electron lens" used in electron microscopes and for electric field focusing of the electron beam in the cathode-ray tube.

Cathode-Ray Tubes

The television picture tube and the precision electron display tube used in oscilloscopes are modern versions of the evacuated tubes used by Crookes and Thompson to study cathode rays. The electrons constituting a "cathode ray" have little mass or inertia, and therefore they can follow rapid variations; their ratio of charge to mass is high, so they are easily deflected and controlled. The energy of high-velocity electrons is easily converted into visible light; therefore, their motion is easily observed. For these reasons the "C-R tube" is a unique information processing device; from our standpoint it is also an ingenious application of the principles of electron motion.

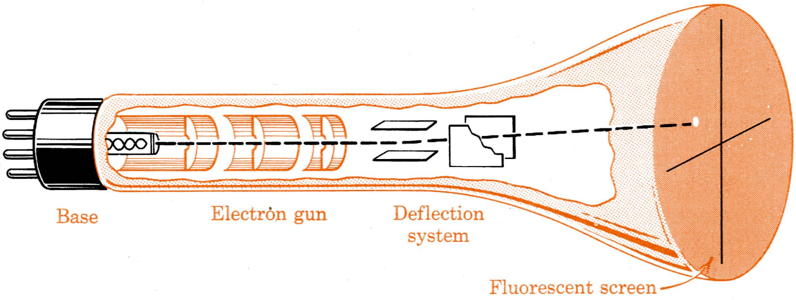
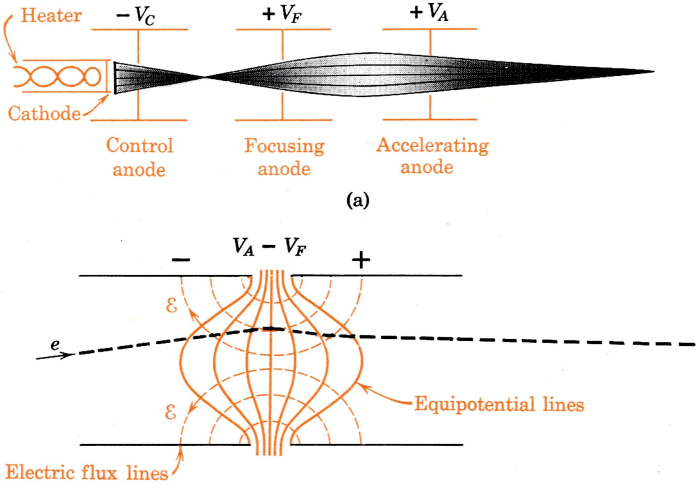


Figure 10.5 The essential components of a cathode-ray tube.

CRT Components

The essential components are shown in Fig. 10.5; an *electron gun* produces a focused beam of electrons, a *deflection system* changes the direction of the beam, and a *fluorescent screen* converts the energy of the beam into visible light.

Electron Gun. Electrons are emitted from the hot cathode (Fig. 10.6a) and pass through a small hole in the cylindrical control electrode; a negative voltage (with respect to the cathode) on this electrode tends to repel the electrons and, therefore, the voltage applied controls the intensity of the beam. Electrons passing the control electrode experience an accelerating force due to the electric field established by the positive voltages V_F and V_A on the focusing and accelerating anodes. The space between these anodes constitutes an electron lens (Fig. 10.6b); the electric flux lines and equipotential lines resulting from the voltage difference $V_A - V_F$ provide a precise focusing effect. A diverging electron is accelerated forward by the field, and at the same time it receives an inward component of velocity that brings it back to the axis of the beam at the screen.



(a)

(b)

Figure 10.6 An elementary electron gun with electric field focusing.

Deflection System. In a television picture tube, the beam is moved across the screen 15,750 times per second, creating a picture consisting of 525 horizontal lines of varying intensity. The deflection of the beam may be achieved by a magnetic field or an electric field. Figure 10.7 shows the beam produced in the electron gun entering the vertical deflection plates of an electric field system.

The beam deflection at the screen can be calculated using our knowledge of electron motion. For the coordinate system of Fig. 10.7, Eq. 10-11 gives an

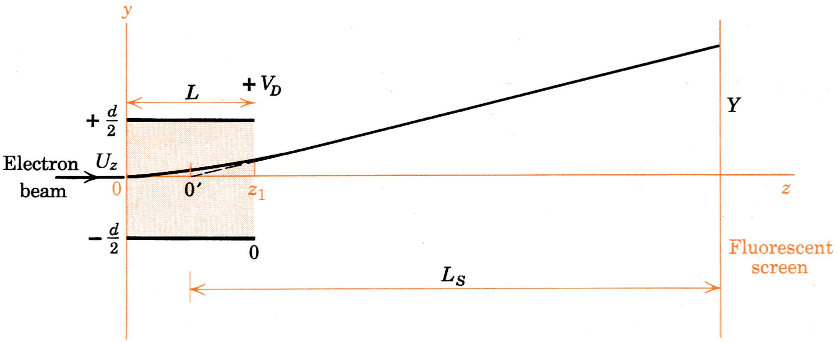


Figure 10.7 An electric field deflection system.

axial velocity

$$U_z = \sqrt{2eV_A/m} \tag{10-16}$$

where V_A is the total accelerating potential. Within the deflecting field, the parabolic path (Eq. 10-7) is defined by

$$y = \frac{eV_D}{2mdU_z^2} z^2 = kz^2 \tag{10-17}$$

The slope of the beam emerging from the deflecting field at $z = z_1 = L$ is

$$\frac{dy}{dz} = 2kz = 2kL \tag{10-18}$$

The equation of the straight-line path followed by the beam to the screen is

$$y - y_1 = \frac{dy}{dz}(z - z_1) = 2kL(z - z_1)$$

where $z_1 = L$ and $y_1 = kz_1^2 = kL^2$. Substituting these values and solving,

$$y = 2kL(z - L) + kL^2 = 2kL\left(z - \frac{L}{2}\right) \tag{10-19}$$

Since for $y = 0$, $z = L/2$, Eq. 10-19 leads to the conclusion that the electron beam appears to follow a straight-line path from a virtual source at 0′. At the screen $z = L_S + L/2$ and (by Eqs. 10-17 and 10-16) the deflection is

$$Y = 2kLL_S = \frac{eV_DLL_S}{mdU_z^2} = \frac{eV_DLL_S}{md} \cdot \frac{m}{2eV_A} = \frac{LL_S}{2dV_A} V_D \tag{10-20}$$

or the vertical deflection at the screen is directly proportional to V_D, the voltage applied to the vertical deflecting plates. A second set of plates provides horizontal deflection. (See Example 3, p. 264.)

Example 3

Determine the *deflection sensitivity* in centimeters of deflection per volt of signal for a C-R tube in which $L = 2$ cm, $L_S = 30$ cm, $d = 0.5$ cm, and the total accelerating voltage is 2 kV.

From Eq. 10-20, the deflection sensitivity is

$$\frac{Y}{V_D} = \frac{LL_S}{2dV_A} = \frac{0.02\ \text{m} \times 0.3\ \text{m}}{2 \times 0.005\ \text{m} \times 2000\ \text{V}} \times 100$$
$$= 0.03\ \text{cm/V}$$

To obtain a reasonable deflection, say, 3 cm, a voltage of 100 V would be necessary. In a practical CRO, amplifiers are provided to obtain reasonable deflections with less than 0.1-V input signals.

Fluorescent Screen. Part of the kinetic energy of the electron beam is converted into luminous energy at the screen. Absorption of kinetic energy results in an immediate *fluorescence* and a subsequent *phosphorescence.* The choice of screen material depends on the application. For laboratory oscilloscopes, a medium persistence phosphor with output concentrated in the green region is desirable; the eye is sensitive to green and the persistence provides a steady image of a repeated pattern. For color television tubes, short persistence phosphors that emit radiation at various wavelengths are available. For radar screens a very long persistence is desirable.

The Cathode-Ray Oscilloscope

The cathode-ray tube (CRT) provides a controlled spot of light whose x and y deflections are directly proportional to the voltages on the horizontal and vertical deflecting plates. The cathode-ray oscilloscope (CRO), consisting of the

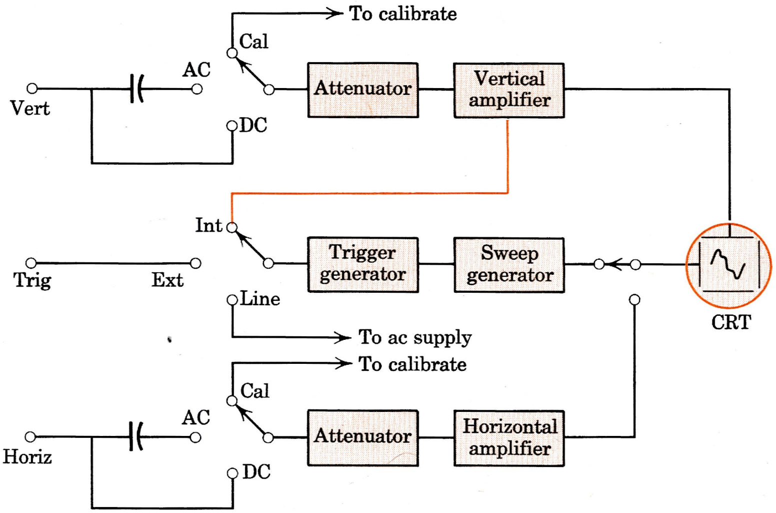

To calibrate
Cal
AC
Vert
DC
Attenuator
Vertical amplifier
Int
Trig
Ext
Line
Trigger generator
Sweep generator
CRT
To ac supply
To calibrate
Cal
AC
Horiz
DC
Attenuator
Horizontal amplifier

Figure 10.8 Basic components of a cathode-ray oscilloscope.

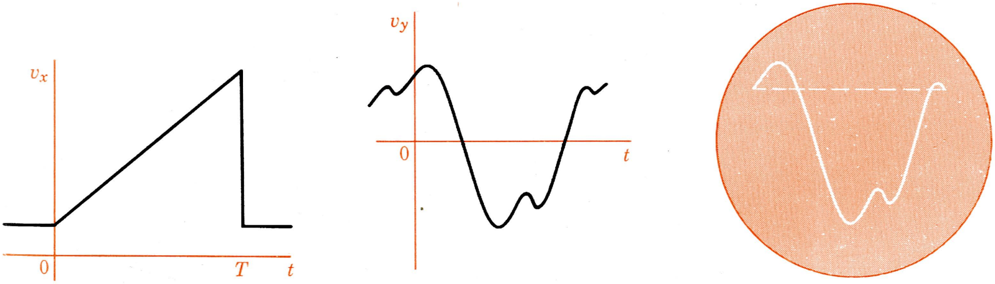
tube and appropriate auxiliary apparatus, is a precise and flexible laboratory instrument. The input characteristics are ideal for measurement purposes and the frequency response may extend to many megahertz.

The block diagram of Fig. 10.8 indicates the essential components. A signal applied to the "Vert" terminal causes a proportional vertical deflection of the spot; the calibration of the *vertical amplifier* can be checked against an internal calibrating signal. The *attenuator* precisely divides large input voltages. The *sweep generator* causes a horizontal deflection of the spot proportional to time; it is *triggered* to start at the left of the screen at a particular instant on the internal vertical signal ("Int"), an external signal ("Ext"), or the ac supply ("Line"). Instead of the sweep generator, the *horizontal amplifier* can be used to cause a deflection proportional to a signal at the "Horiz" terminal. An *intensity control* varies the accelerating potentials in the electron gun, and a *focus control* determines the potentials on the focusing electrodes.

Voltage Measurement. An illuminated scale dividing the screen into 1-cm divisions permits use of the CRO as a voltmeter. With the vertical amplifier sensitivity set at 0.1 V/cm, say, a displacement of 2.5 cm indicates a voltage of 0.25 V. Currents can be determined by measuring the voltage across a known resistance.

Time Measurement. A calibrated sweep generator permits time measurement. With the sweep generator set at 5 ms/cm, two events separated on the screen by 2 cm are separated in time by 10 ms or 0.01 s. By measuring the period of a wave, the frequency can be determined by calculation.

Waveform Display. A special property of the CRO is its ability to display high-frequency or short-duration waveforms. If voltages varying with time are applied to vertical (y) and horizontal (x) input terminals, a pattern is traced out on the screen; if the voltages are periodic and one period is an exact multiple of the other, a stationary pattern can be obtained. A sawtooth wave from the sweep generator (Fig. 10.9a) applied to the x-deflection plates provides an x-axis deflection directly proportional to time. If a signal voltage wave is applied to the vertical-deflection plates, the projection of the beam on the y axis is directly



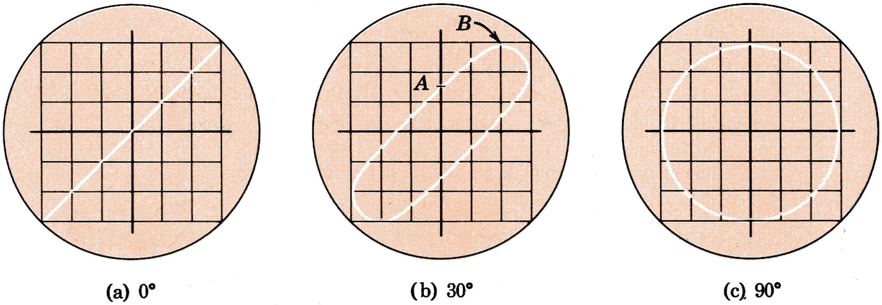
(a) Saw-tooth sweep (b) Signal wave-form (c) Oscilloscope pattern

Figure 10.9 Display of a repeated waveform on an oscilloscope.

proportional to the amplitude of this signal. If both voltages are applied simultaneously, the pattern displayed on the screen is the signal as a function of time (Fig. 10.9c). A *blanking circuit* turns off the electron beam at the end of the sweep so the return trace is not visible. Nonrepetitive voltages are made more visible by using a long persistence fluorescent material, or a high-speed camera can be used for a permanent record.

X-Y Plotting. The relation between two periodic variables can be displayed by applying a voltage proportional to x to the horizontal amplifier and one proportional to y to the vertical amplifier. The characteristics of diodes or transistors are quickly displayed in this way. The hysteresis loop of a magnetic material can be displayed by connecting the induced emf (proportional to B) to the vertical amplifier and an iR drop (proportional to H) to the horizontal amplifier. Since the two amplifiers usually have a common internal ground, some care is necessary in arranging the circuits.

Phase-Difference Measurement. If two sinusoids of the same frequency are connected to the X and Y terminals, the phase difference is revealed by the resulting pattern (Fig. 10.10). For applied voltages $v_x = V_x \cos \omega t$ and $v_y = V_y \cos(\omega t + \theta)$, it can be shown that the phase difference is



(a) 0° (b) 30° (c) 90°

Figure 10.10 Phase difference as revealed by CRO patterns.

$$\theta = \sin^{-1}\frac{A}{B} \tag{10-21}$$

where A is the y deflection when the x deflection is zero and B is the maximum y deflection. The parameters of the circuit are useful in determining whether the angle is leading or lagging.

Frequency Comparison. When the frequency of the sinusoid applied to one input is an exact multiple of the frequency of the other input, a stationary pattern is obtained. For a 1:1 ratio, the so-called Lissajous patterns are similar to those in Fig. 10.10. For 1:2 and 2:3 ratios, the Lissajous figures might be as shown in Fig. 10.11. For a stationary pattern, the ratio of the frequencies is exactly equal to the ratio of the number of tangencies to the enclosing rectangle. Patterns can be predicted by plotting x and y deflections from the two signals at corresponding instants of time.

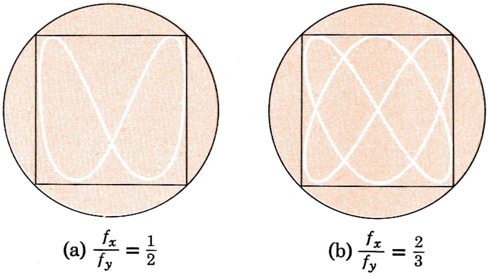

(a) $\frac{f_x}{f_y} = \frac{1}{2}$ (b) $\frac{f_x}{f_y} = \frac{2}{3}$

Figure 10.11 Lissajous figures for comparing the frequencies of two sinusoids.

Vacuum Diodes

The cathode-ray tube can be analyzed in terms of the effect of electric and magnetic fields on a single electron in space. As a next step we now consider the case of an evacuated space containing many electrons, as in an ordinary vacuum tube. First, we look at the process for obtaining large quantities of electrons, then we consider the effect of the resulting *space charge*, and on this basis we determine the behavior of a two-electrode vacuum tube in terms of its v-i characteristics.

Electron Emission

The Bohr model of the atom is satisfactory for describing how free electrons can be obtained in space. As you recall, starting with the hydrogen atom consisting of a single proton and a single orbital electron, models of more complex atoms are built up by adding protons and neutrons to the nucleus and electrons in orbital groups or shells. In a systematic way, shells are filled and new shells started. The chemical properties of an element are determined by the *valence* electrons in the outer shell. Good electrical conductors like copper and silver have only one electron in the outer shell.

Only certain orbits are allowed, and atoms are stable only when the orbital electrons have certain discrete energy levels. Transfer of an electron from an orbit corresponding to energy W_1 to an orbit corresponding to a lower energy W_2 results in the radiation of a *quantum* of electromagnetic energy of frequency f given by

$$W_1 - W_2 = hf \tag{10-22}$$

where h is Planck's constant $= 6.626 \times 10^{-34}$ J-s.

The energy possessed by an orbital electron consists of the kinetic energy of motion in the orbit and the potential energy of position with respect to the positive ion representing all the rest of the neutral atom. If other atoms are close (as in a solid), the energy of an electron is affected by the charge distribution of the neighboring atoms. In a crystalline solid, there is an orderly arrangement of atoms and the permissible electron energies are grouped into *energy bands.* Between the permissible bands there may be ranges of energy called *forbidden bands.*

For an electron to exist in space it must possess the energy corresponding to motion from its normal orbit out to an infinite distance; the energy required to move an electron against the attractive force of the net positive charge left behind is the *surface barrier energy* W_B. Within a metal at absolute zero temperature, electrons possess energies varying from zero to a maximum value W_M. The minimum amount of work that must be done on an electron before it is able to escape from the surface of a metal is the *work function* W_W where

$$W_W = W_B - W_M \tag{10-23}$$

For copper, $W_W = 4.1$ eV, while for cesium $W_W = 1.8$ eV.

The energy required for electron emission may be obtained in various ways. The "beta rays" given off spontaneously by *radioactive* materials (along with alpha and gamma rays) are emitted electrons. In *photoelectric* emission, the energy of a quantum of electromagnetic energy is absorbed by an electron. In *high-field* emission, the potential energy of an intense electric field causes emission. In *secondary* emission, a fast-moving electron transfers its kinetic energy to one or more electrons in a solid surface. All these processes have possible applications, but the most important process is *thermionic* emission in which thermal energy is added by heating a solid conductor.

The temperature of an object is a measure of the kinetic energy stored in the motion of the constituent molecules, atoms, and electrons. The energies of the individual constituents vary widely, but an average energy corresponding to temperature T can be expressed as kT, where $k = 8.62 \times 10^{-5}$ eV/K is the Boltzmann constant. At a temperature above absolute zero, the distribution of electron energy in a metal is modified and some electrons possess energies appreciably above W_M. Statistical analysis shows that the probability of an electron receiving sufficient energy to be emitted is proportional to $e^{-W_W/kT}$. At high temperatures, many electrons possess energies greater than W_B and emission current densities of the order of 1 ampere per square centimeter are practical. Commercial cathodes make use of special materials that combine low work function with high melting point.

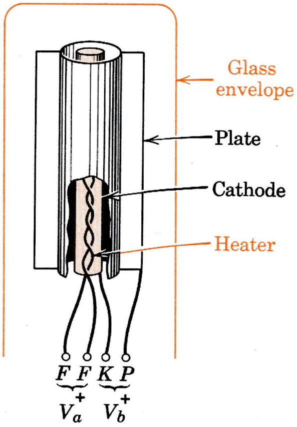


Figure 10.12 Construction of a typical vacuum diode.

Space Charge

The construction of a typical vacuum diode is shown in Fig. 10.12. A cylindrical *anode* or *plate* is mounted in an evacuated glass envelope. A cylindrical *cathode* is heated by a high-resistance *heater* supplied by a voltage V_a. The plate is maintained at a positive potential V_b with respect to the cathode. Negative electrons emitted from the cathode are drawn to the positive plate; after flowing through an external load circuit, the electrons return to the cathode.

We are interested in the *v-i* characteristics that permit this electronic device to perform the important functions of detection and rectification. To simplify the analysis, let us consider the idealized diode of Fig. 10.13a in which the electrodes are parallel planes, and only a few electrons are present in the evacuated space. Superimposed on the sketch of Fig. 10.13b is a graph of potential v in the cathode-plate space. There is a corresponding uniform negative electric field intensity $\mathcal{E}$. An electron emitted from the cathode with zero velocity experiences an accelerating force equal to $-e\mathcal{E}$, and its velocity u at any point is proportional to $\sqrt{v}$ (Eq. 10-11).

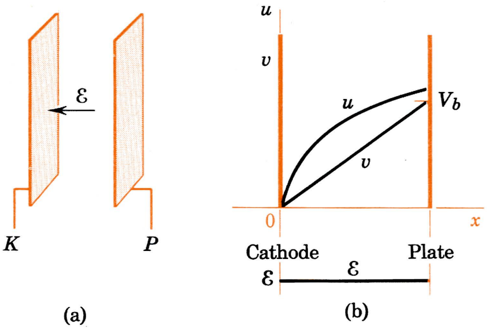


Figure 10.13 Potential, field, and velocity distributions in an idealized vacuum diode.

If the cathode temperature is raised, more and more electrons are emitted, some with appreciable initial velocity, and there are many in the space at all times. (A current of 10 mA corresponds to 6×10^{16} electrons/s.) Repulsive forces cause the electrons close to the cathode to accelerate more slowly and those close to the plate to accelerate more rapidly; the final velocity is determined by V_b since the kinetic energy gained is $\frac{1}{2}(mu_b{}^2 - mU_0{}^2) = eV_b$. Up to a point, the plate current in electrons per second is just equal to the rate of emission. A further increase in cathode temperature increases the rate of emission, but some electrons are repelled by the space charge and "fall" back into the cathode; the plate current is said to be *space-charge limited.*

The conditions existing in the cathode-plate space are described by the graphs in Fig. 10.14. The governing equations are

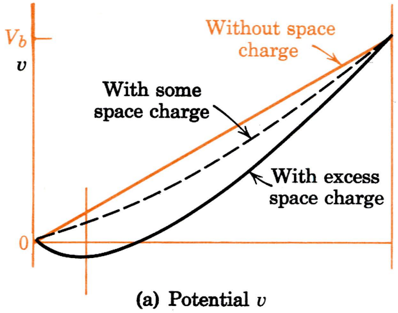
$$u = \sqrt{U_0{}^2 + 2ev/m} \tag{10-24}$$

$$\mathcal{E} = -\frac{dv}{dx} \tag{10-25}$$

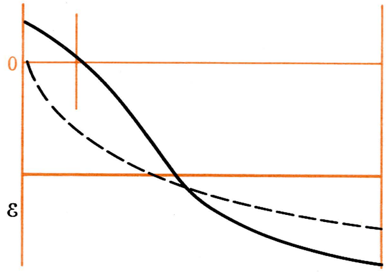
$$\rho u = J \tag{10-26}$$

$$\rho = \epsilon \frac{d\mathcal{E}}{dx} \tag{10-27}$$

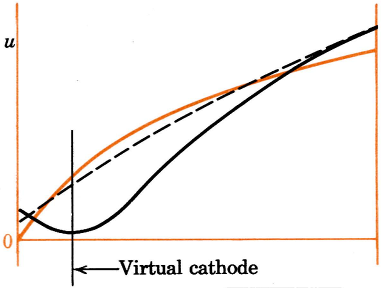
Equations 10-24 and 10-25 are familiar. Equation 10-26 expresses the fact that the current density J (A/m^2) is the product of charge density ρ (rho) (C/m^3) and velocity u (m/s); the current density is constant across the space since the electrons passing one intermediate plane must pass any other plane. Equation



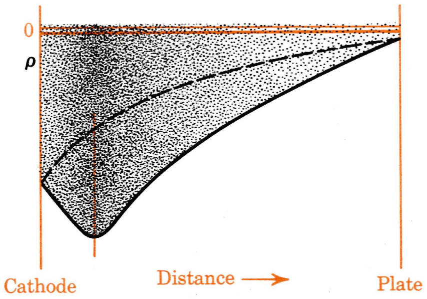
(a) Potential v



(b) Electric field $\mathcal{E} = -\frac{dv}{dx}$



(c) Velocity $u = \sqrt{U_O{}^2 + 2ev/m}$



(d) Charge density $\rho = \frac{J}{u}$

Figure 10.14 Conditions in a diode.

10-27 is based on Gauss' law, which says that the net electric flux passing through any closed surface is equal to the total charge enclosed by the surface.†

Vacuum tubes are always operated with cathode temperatures high enough to provide excess emission, and the actual voltage distribution is as shown by the solid black line in Fig. 10.14a. The electric field distribution ($\mathcal{E} = -dv/dx$) is as shown in Fig. 10.14b; the positive field at the cathode repels all but the most energetic emitted electrons. As shown in Fig. 10.14c, even the fast-moving electrons are slowed down by the negative space charge, reach a minimum velocity at the point of minimum voltage (Eq. 10-24), and then are accelerated. The region of minimum velocity is a region of maximum charge density (Fig. 10.14d) since current density is constant (Eq. 10-26). The effect is as if electrons were being emitted at low velocity from a region outside the cathode, so we say that a *virtual cathode* exists there.

Vacuum Diode Characteristics

To derive the *v-i* characteristic for a vacuum diode, we must solve the four governing equations (Eqs. 10-24 to 10-27) simultaneously. If initial velocity U_O is neglected or if V_b (at $x = b$) is measured with respect to the potential of the virtual cathode (at $x = 0$), the solutions are:

$$v = V_b\left(\frac{x}{b}\right)^{4/3} \qquad \mathcal{E} = \mathcal{E}_b\left(\frac{x}{b}\right)^{1/3}$$

$$u = u_b\left(\frac{x}{b}\right)^{2/3} \qquad \rho = \rho_b\left(\frac{x}{b}\right)^{-2/3}$$

---

† $\psi = \oint D\,dA = \epsilon \oint \mathcal{E}\,dA = \int dq = \int \rho A\,dx$
where ψ = electric flux (psi),
D = electric flux density = ψ/A,
ϵ = permittivity (epsilon) = $D/\mathcal{E}$,
$A\,dx$ = element of volume (Fig. 10.15).
Solving this equation for ρ yields Eq. 10-27.

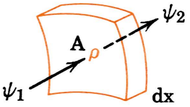


Figure 10.15 $d\psi = \rho A\,dx$.

(As one check on these solutions, note that $J = \rho u = \rho_b u_b =$ constant.) For space-charge limited operation, the charge density near the plate is

$$\rho_b = \epsilon \frac{d\mathcal{E}}{dx}\bigg|_{x=b} = -\epsilon \frac{d^2 v}{dx^2}\bigg|_{x=b} = -\frac{4\epsilon}{9b^2} V_b = -k_1 V_b$$

By Eq. 10-24 for $U_0 = 0$, the velocity near the plate is

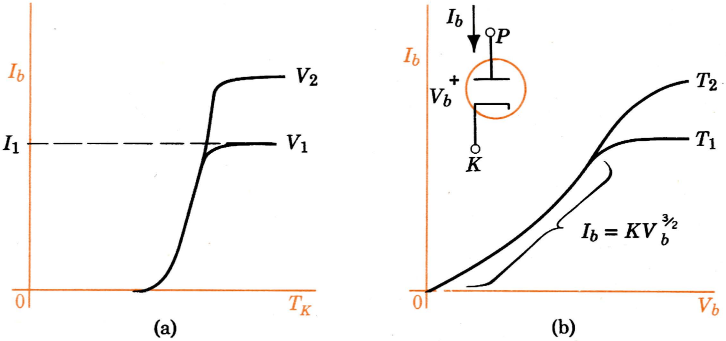
$$u_b = \left(\frac{2e}{m}\right)^{1/2} V_b^{\,1/2} = k_2 V_b^{\,1/2}$$

The *plate current* (from the anode to the cathode in Fig. 10.16) is

$$I_b = -J_b A = k_1 k_2 A V_b V_b^{\,1/2} = K V_b^{\,3/2} \tag{10-28}$$

Experimental observations on diodes with parallel-plane electrodes agree closely with the "$\frac{3}{2}$ power law" predicted by Eq. 10-28. Experimental results for practical tubes with electrodes in the form of concentric cylinders (Fig. 10.12) also agree quite well with Eq. 10-28, but design calculations are usually based on graphical data supplied by the manufacturer.

The effects of cathode temperature and plate voltage on plate current are displayed in Fig. 10.16. The electron emission (proportional to $e^{-W_w/kT}$) rises rapidly with cathode temperature; if the plate voltage is sufficiently high, all the electrons are drawn across the space and the plate current rises correspondingly (Fig. 10.16a). Because of space charge, however, the plate current is limited to I_1 for a given V_1 (Eq. 10-28), and a further increase in cathode temperature does not result in increased current. Raising the plate voltage to V_2 raises the possible current. As shown in Fig. 10.16b, for a given cathode temperature T_1, plate current increases with plate voltage according to the $\frac{3}{2}$ power law until all the emitted electrons are being drawn across the space. Operation under this condition is said to be "temperature limited"; a further increase in current requires an increase in cathode temperature. Vacuum tubes



(a) (b)

Figure 10.16 The effects of cathode temperature and plate voltage on plate current in a vacuum diode.

are ordinarily operated in the "space-charge limited" region in which $I_b \cong KV_b^{3/2}$.

Example 4

A vacuum diode is operated with a heater voltage $V_a = 6$ V and a plate voltage $V_b = 100$ V; the plate current I_b is 12 mA. Increasing V_a to 6.5 V produces no change in I_b. Predict the value of I_b for $V_a = 7$ V and $V_b = 25$ V.

The fact that increasing V_a, and therefore cathode temperature, produces no change in I_b indicates that the tube is operating in the space-charge limited region. Therefore, the increase in V_a above 6 V is not significant and

$$\frac{I_b'}{I_b} = \frac{KV_b'^{3/2}}{KV_b^{3/2}} = \left(\frac{V_b'}{V_b}\right)^{3/2}$$

Solving,

$$I_b' = I_b\left(\frac{V_b'}{V_b}\right)^{3/2} = 12\left(\frac{25}{100}\right)^{3/2} = 12 \times \tfrac{1}{8} = 1.5 \text{ mA}$$

Gas Diodes

The effect of large numbers of negative electrons in the cathode-plate space is to limit the current. What is the effect of positive ions in this space? If a small amount of a gas, such as argon or mercury vapor, is introduced, the effect at low plate voltages is small; electrons crossing the space may collide with one or more atoms of gas, give up some kinetic energy, and then proceed to the plate. If, however, the electron imparts sufficient energy to ionize the gas atoms, there are two significant effects. First, ionization produces a free electron that adds to the current flow. Second, the positive ion moves toward the cathode and tends to neutralize the negative space charge. The resulting increased flow of electrons increases the probability of ionization, decreases the space charge further, and increases the current still more. The space charge becomes positive due to the presence of many slow-moving ions and the current is limited solely by cathode emission. The resulting gas diode characteristic is shown in Fig. 10.17; in such a tube containing mercury vapor (ionization potential = 10.4 V), a large current can flow with just a small applied voltage and, therefore, with just a small power loss.

Photo-Emissive Diodes

In thermionic diodes, emission is obtained by heating the cathode. A diode in which electrons are obtained by photoelectric emission is called a "phototube." In one type, the cathode is a half-cylinder with the inside surface coated with a material sensitive to visible light. The emitted electrons are collected at a slender anode mounted in front. As shown in Fig. 10.18, for a given luminous flux the current increases with voltage until all the electrons are drawn to the

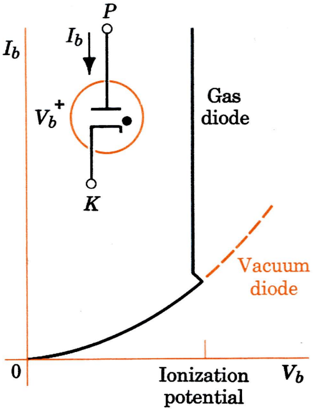


Figure 10.17 Gas diode characteristic.

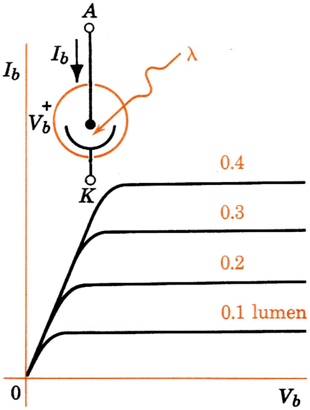


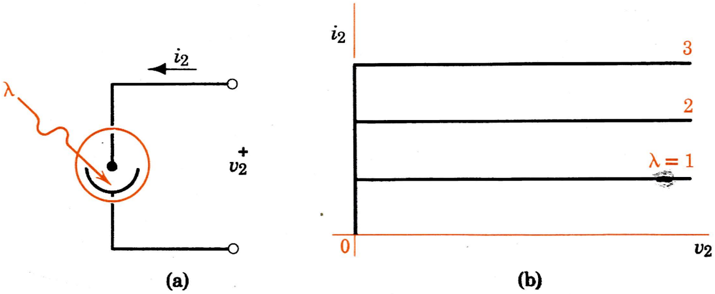
Figure 10.18 Phototube characteristics.

anode. For voltages above a certain value, the current is proportional to the luminous flux. For example, such tubes are used in motion picture projectors to convert illumination from the sound track into electric currents for amplification and reproduction.

Two-Port Electronic Devices

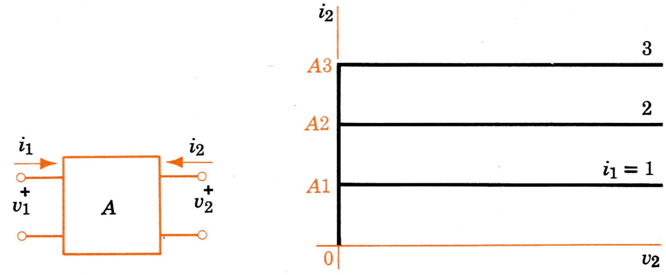
The vacuum diode exhibits a useful control property: it permits appreciable current flow under forward-bias voltage and tends to prevent current flow when the voltage is reversed. Such a one-port device is sometimes called a "switch." Electrically speaking, a phototube is also a diode, but the current flow for a given voltage is influenced by the light signals introduced at a second port. In idealized form, the characteristic curves of Fig. 10.19b represent the phototube as a two-port *transducer.*

An *amplifier* is an important example of an electronic two-port. As a result of voltage or current signals applied to the input, larger voltage or current signals are available at the output. In the idealized current amplifier of Fig. 10.20, the



(a) (b)

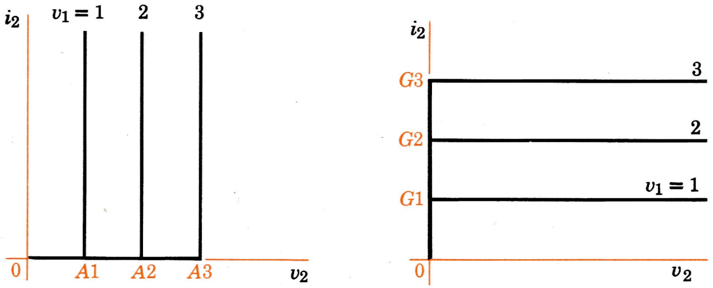
Figure 10.19 A phototube as a transducer.



(a) (b)

Figure 10.20 An idealized current amplifier.

output current for any output voltage is A times as great as the input current.† In the idealized voltage amplifier of Fig. 10.21a, the output voltage is A times the input voltage for any output current. In the idealized transducer of Fig. 10.21b, the output current is G times the input voltage; such a device can be converted into a voltage amplifier by sending the output current through a load resistance.



(a) Voltage amplifier **(b) General transducer**

Figure 10.21 Idealized device characteristics.

Vacuum Triode

Perhaps Lee deForest did not have Fig. 10.21a in mind in 1906, but he did invent a device whose characteristics approach those shown. One of the first radio engineers, deForest had developed his own wireless system using vacuum diode detectors. In various experiments to improve reception he used tinfoil wrapped around the tube, a second plate on the other side of the filament, a perforated platinum sheet between the filament and the plate, and finally a folded wire "grid," which worked best of all (Fig. 10.22). The resulting triode‡ changed the course of electronics because of its ability to amplify signals. The conventional symbol for a triode is shown in Fig. 10.23a; the heater is usually omitted from circuit diagrams.

---

† A truly "ideal" amplifier would function just as well for negative currents and voltages, and the characteristic curves would extend into all four quadrants.

‡ The patents for this simple device, granted after 19 years of stormy litigation, were among the most valuable ever issued: Paul G. Watson, "deForest's Early Audions," *Electric World*, September 1965.

The operation of the triode can be explained in terms of the effect of the grid on the potential distribution in the cathode-plate space. Assuming space-charge limited operation, the potential distribution in a diode is as shown by the solid line in Fig. 10.23b. Because of the excess electrons emitted from the cathode, a space charge exists and tends to repel the emitted electrons back to the cathode; only the more energetic electrons pass the potential barrier and are drawn to the plate.

If an open-wire grid is placed in the space, it tends to assume the potential of that place but does not impede electron flow because of the low ratio of wire width to grid opening. If a positive potential $+v_c$ is placed on the grid, the potential distribution is modified as shown. The effect can be described as a lowering of the potential barrier or as a reduction in the field strength (the slope of the potential curve). In either case, the result is an increased net current from the cathode. Some of the electrons are attracted to the positive grid, but most of the current goes through the grid to the plate. If a negative potential $-v_c$ is placed on the grid, the potential barrier is raised and the plate current is reduced. The negative grid repels electrons and there is no grid current.

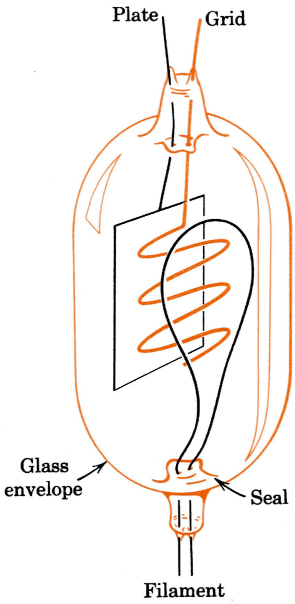
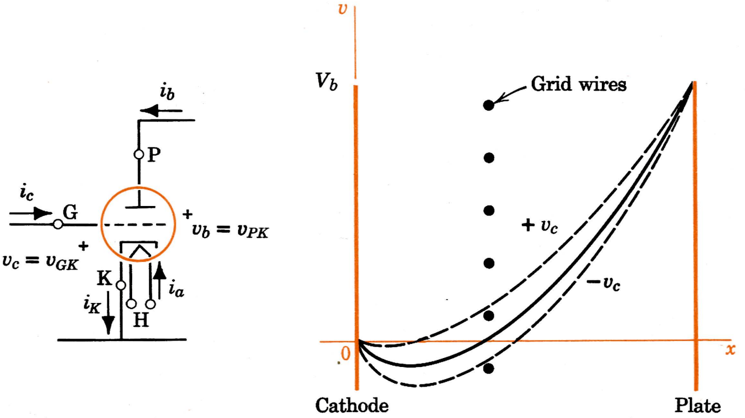


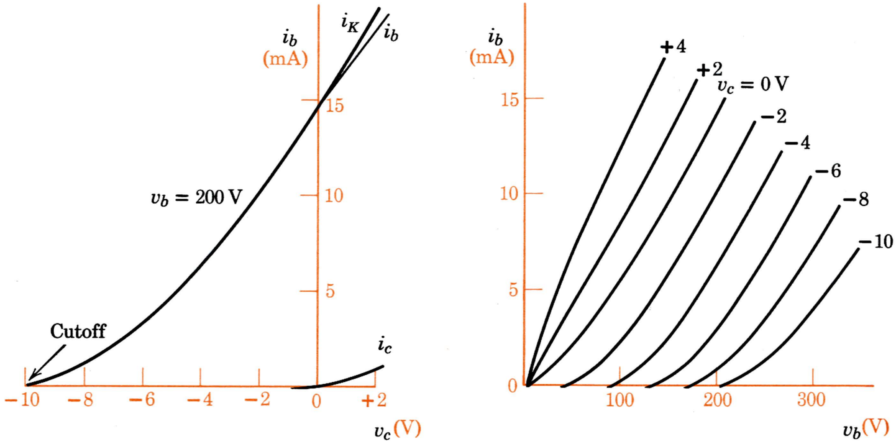
Figure 10.22 An early form of deForest triode.



(a) Notation (b) Potential distribution

Figure 10.23 The vacuum triode.

For a given plate voltage, the variation of current with grid voltage is as shown in Fig. 10.24a. For positive grid voltages, the cathode current is divided between plate and grid. At $v_c = 0$, a small grid current i_c flows due to electrons striking the grid wires. A small negative potential on the grid repels even the most energetic electrons and reduces this current to zero. As the grid is made more negative, the plate current i_b is reduced and becomes zero at a grid voltage called the *cutoff* value. A family of such curves drawn for various values of plate voltage constitutes the *transfer* characteristics i_b versus v_c.



(a) Grid (b) Plate

Figure 10.24 Typical triode characteristics.

An alternative method of displaying triode characteristics is shown in Fig. 10.24b. Here plate current i_b is shown as a function of plate voltage v_b for various values of grid voltage v_c. The cathode current can be approximated by the theoretical relation

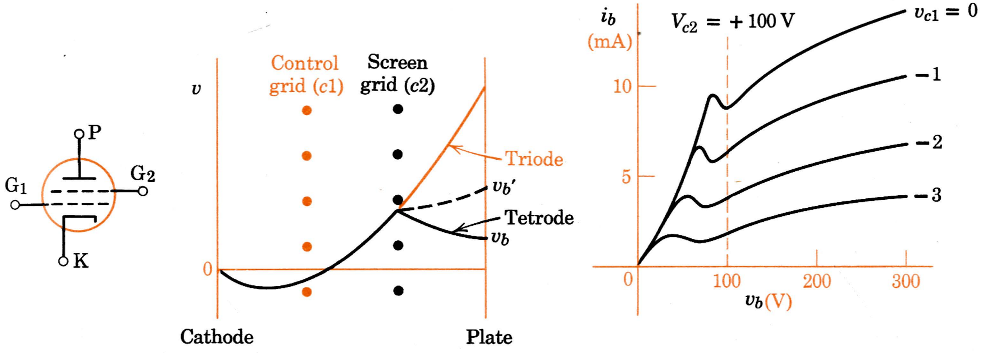
$$i_k = K(\mu v_c + v_b)^{3/2} \tag{10-29}$$

For negative grid operation, $i_k = i_b$ and the curves resemble the $\frac{3}{2}$ power curve of the diode. The grid, being closer to the cathode, is more effective than the plate by the factor μ in Eq. 10-29. Note the resemblance between these curves and those of the idealized amplifier of Fig. 10.21a. With a variation in grid voltage, an input signal, there is a variation in plate voltage, an output signal. Also note that if the grid is negative, no grid current flows and the power input is zero. This is a desirable feature in an amplifier. In fact, one of the great virtues of the triode is its ability to control an appreciable output power with practically no input signal power.

Tetrodes and Pentodes

The relations shown in Fig. 10.24 are called *static* characteristics because they are obtained from dc measurements. In radiofrequency amplification, the signals may consist of voltage variations at megahertz frequencies. Since $i = C\,dv/dt$, at these frequencies the very small capacitances between grid and cathode, grid and plate, and plate and cathode draw appreciable current. These *interelectrode capacitances* may be the limiting factors in the high-frequency performance of the tube. The grid-plate capacitance, only a few picofarads, is particularly annoying because it not only decreases the possible amplification but it also provides a path whereby some of the output signal is fed back into the input and unwanted oscillations may occur.

One way to reduce the grid-plate capacitance is to insert a second grid (Fig. 10.25a, b) that provides an electrostatic shield between *control grid* ($c1$) and plate.



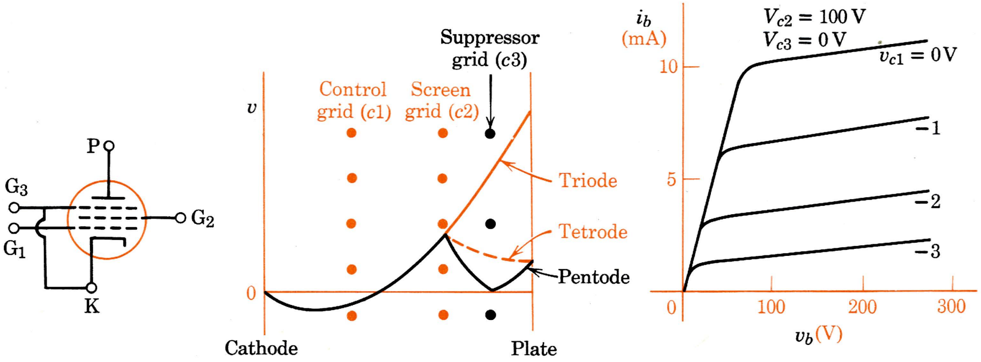
(a) Symbol (b) Potential distribution (c) Plate characteristics

Figure 10.25 The vacuum tetrode.

This *screen grid* ($c2$) is maintained at a fixed positive potential with respect to the cathode. It draws a small current, but the bulk of the current flows through the screen grid to the plate. One result of the screen grid, an advantage, is that variations in plate voltage (for example, from v_b to v_b') have little effect on the potential distribution near the cathode and, therefore, have little effect on the plate current (Fig. 10.25b). The relative effect of the grid voltage is increased, i.e., the value of μ in Eq. 10-29, and therefore the voltage amplification capability.

A second result, a disadvantage in amplification, is that the plate characteristics become highly nonlinear. The dips in the plate current curves of Fig. 10.25c can be explained in terms of the potential distribution between the screen grid and the plate. In any vacuum tube, electrons accelerated by more than a few volts gain sufficient kinetic energy to knock electrons out of the plate in the process called *secondary emission.* In a diode or triode, the electric field near the plate forces the secondary electrons back to the plate. In a tetrode at low plate voltages (v_b in Fig. 10.25b), the electric field is reversed, secondary electrons are attracted to the more positive screen grid, and the plate current is effectively reduced.

The kinks in the plate characteristics of a tetrode can be removed by inserting another grid between the screen grid and the plate (Fig. 10.26a). Normally connected to the cathode, and therefore at zero potential, this *suppressor grid* ($c3$) depresses the potential distribution curve (Fig. 10.26b) and restores the negative electric field in front of the plate. Secondary emission from the plate is suppressed and the plate characteristics, shown in Fig. 10.26c, approach those of the idealized amplifier (Fig. 10.21b). Another method of obtaining the desired depression of the potential distribution in front of the plate is to use the space charge produced by a concentrated beam of electrons. Such a *beam power* tube has characteristics that even more closely approach those of the idealized transducer.



(a) Symbol (b) Potential distribution (c) Plate characteristics

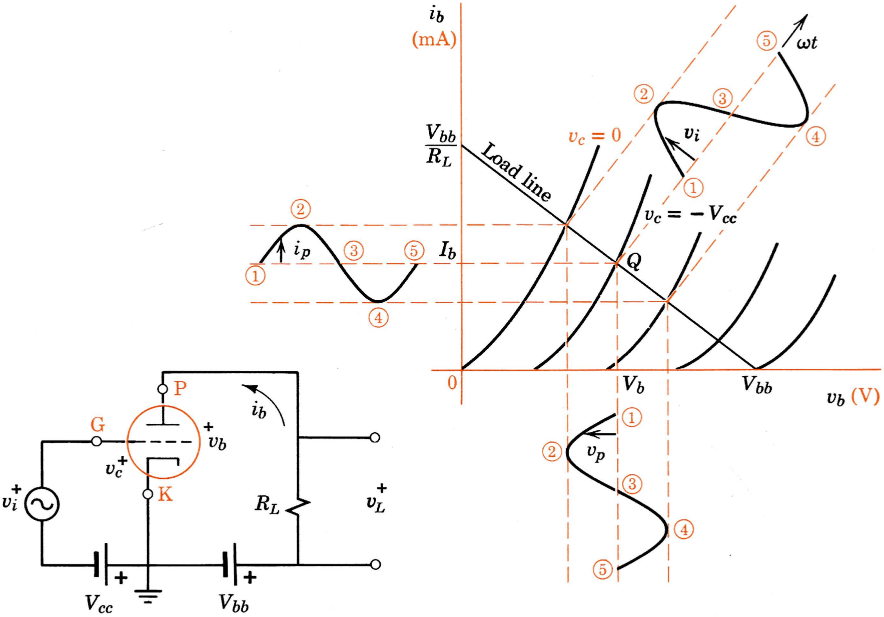
Figure 10.26 The vacuum pentode.

The evolution of multielectrode vacuum tubes from the simple diode is a good illustration of skilled invention guided by basic knowledge and stimulated by practical need. The diode acts as an efficient switch. The triode and pentode provide great amplification. The cathode-ray tube is a sensitive instrument. The linear accelerator imparts enormous energies to tiny electrons. The key ingredient in each is the engineer's ability to provide electric fields that precisely and predictably control the motion of charged particles.

An Elementary Amplifier

An electronic amplifier consists of a device having the proper external characteristics, power supplies to place the device in the proper operating condition, a means for introducing the input signal, and a means for extracting the output signal. Practical aspects of amplifier design and operation are presented in Chapters 16 and 17 and applied to small-signal and feedback amplifiers in Chapters 18 and 19. To conclude this discussion of principles, let us consider the analysis of an elementary form of electronic amplifier by using the graphical approach of Chapter 2.

In Fig. 10.27, batteries V_{cc} and V_{bb} supply power to the triode tube whose characteristics are shown. The input signal is applied in series with the negative grid-bias voltage of magnitude V_{cc}. It is convenient to consider a sinusoidal input voltage v_i since any periodic wave can be represented by a series of sinusoids of various frequencies and amplitudes. The plate supply voltage V_{bb} maintains the plate at a positive potential with respect to the cathode. We anticipate that a signal voltage v_i will cause a variation in grid-to-cathode voltage v_c, which will produce a variation in plate current i_b, developing, in turn, an amplified output voltage across the load resistance R_L.



(a) Circuit diagram **(b)** Operation

Figure 10.27 An elementary triode voltage amplifier.

By Kirchhoff's voltage law around the output loop (see Table 10-1 on p. 280),

$$v_b = V_{bb} - i_b R_L \tag{10-30}$$

This is the equation of the *load line,* a straight line (Fig. 10.27b) with intercepts V_{bb} (for $i_b = 0$) and V_{bb}/R_L (for $v_b = 0$), and a slope of $-1/R_L$. Intersections of

Table 10-1 Vacuum-Tube Nomenclature

	Grid-cathode voltage	Plate-cathode voltage	Plate current
Instantaneous total value	v_c	v_b	i_b
Instantaneous signal component	v_g	v_p	i_p
Quiescent or dc value	V_c	V_b	I_b
Effective (rms) value of signal	V_g	V_p	I_p
Supply voltage (magnitude)	V_{cc}	V_{bb}	

the load line with the characteristic curves of the nonlinear triode represent graphical solutions of Eq. 10-30 and define the operating values of v_b, i_b, and v_c.

By Kirchhoff's law around the input loop,

$$v_c = -V_{cc} + v_i \tag{10-31}$$

With no signal, $v_i = 0$ and $v_c = -V_{cc}$. The intersection of the load line with the characteristic curve for $v_c = -V_{cc}$ locates Q, the *quiescent point.* The quiescent plate current is labeled I_b and the quiescent plate voltage is V_b.

With a sinusoidal signal applied, the grid voltage takes successive values corresponding to points 1, 2, 3, 4, 5 on the waveform. The corresponding values of plate voltage and current are obtained graphically from the load line. The signal output v_o is the sinusoidal component of the voltage v_L across R_L, or

$$v_o = v_p = i_p R_L \tag{10-32}$$

where v_p and i_p are instantaneous values of the sinusoidal components of plate voltage and current as shown in Fig. 10.27b. The voltage amplification, or voltage *gain* of the amplifier, is

$$A_V = \frac{v_o}{v_i} = \frac{V_{om}}{V_{im}} = \frac{V_o}{V_i} \tag{10-33}$$

depending on whether the voltage ratio is expressed in terms of instantaneous, maximum, or rms values. Numerical calculations for a typical triode are shown in Example 5.

Future of the Electron Tube

In the preceding discussions, it was assumed that the period of a voltage or current cycle was long compared to the *transit time* required for an electron to go from cathode to anode. If the frequency of the signal is so high that the electric field experienced by an electron changes significantly during the time of transit, ordinary vacuum tubes cannot be used. For the generation and amplification of *microwaves,* at frequencies above 1000 MHz, there are special vacuum tubes such as *klystrons, traveling-wave tubes,* and *magnetrons* that utilize the

Example 5

The triode whose characteristics are given in Fig. 10.28 is used in the circuit of Fig. 10.27a where $V_{cc} = 4$ V, $V_{bb} = 290$ V, and $R_L = 20$ kΩ. Predict the output voltage and voltage gain for $v_i = 4 \sin \omega t$ V.

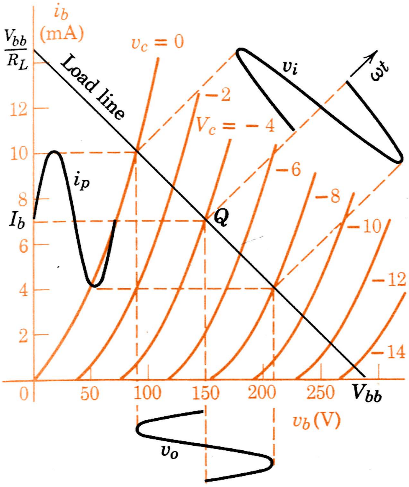


Figure 10.28 Triode amplifier performance.

The intercepts of the load line are determined to be $V_{bb} = 290$ V and $V_{bb}/R_L = 290$ V/20 kΩ = 14.5 mA, and the load line is drawn. The intersection of the load line with the $v_c = -4$ V curve indicates a Q point at $I_b = 7$ mA and $V_b = 150$ V.

For $\omega t = \pi/2$, the signal voltage is $v_i = +4$ V, the grid voltage $v_c = -V_{cc} + v_i = -4 + 4 = 0$, and the corresponding plate current is $i_b = 10$ mA.

For $\omega t = 3\pi/2$, $v_i = -4$ V, $v_c = -4 - 4 = -8$ V, and the corresponding $i_b = 4$ mA.

A signal voltage swing of ±4 V causes a plate current swing of ±3 mA; in terms of peak values, the corresponding output signal is (by Eq. 4-32)

$$V_{om} = I_p R_L = 3 \text{ mA} \times 20 \text{ k}\Omega = 60 \text{ V}$$

This value could also be obtained graphically as a change in v_b from 210 to 150 V. The voltage gain is

$$A_V = \frac{V_{om}}{V_{im}} = \frac{60}{4} = 15$$

transit time in their operation.† In each of these devices a stream of electrons is subjected to electric and magnetic fields precisely arranged so that an exchange of energy occurs between the signal and the electrons. The net result is that power supplied by a dc source is converted to ac power at the signal frequency and amplification is obtained. High operating efficiencies are possible and these microwave tubes find application in television transmission, satellite communication, industrial heating, and home cooking.

The future of the electron tube does not lie in the information processing function that it revolutionized; the processing of analog and digital information at low levels has been taken over by the semiconductor devices to be discussed in the next three chapters. The principal applications of electron tubes, with annual sales of a billion dollars, will be in power conversion (microwave tubes), in light conversion (phototubes), in display (picture tubes), and in power generation at high frequencies (>100 MHz) and high power (>100 W). The concepts presented in this chapter provide only a basis for further study.

---

† See Chapter 19 of F. E. Terman: *Electronic and Radio Engineering*, 4th ed., McGraw-Hill Book Co., New York, 1955.

SUMMARY

◆ Individual charged particles in electric and magnetic fields obey Newton's laws. For electrons (of primary interest here),

$$\mathbf{f} = -e(\mathcal{E} + \mathbf{u} \times \mathbf{B})$$

◆ In any electric field, KE gained = PE lost = $W_{ab} = eV_{ab}$.
For an electron starting from rest (and for $V < 4$ kV),

$$u = \sqrt{2(e/m)V} = 5.93 \times 10^5 \sqrt{V}\ \text{m/s}$$

In a uniform electric field where $\mathcal{E}_x = \mathcal{E}$, $f_x = -e\mathcal{E}$, and

$$a_x = -\frac{e\mathcal{E}}{m} \qquad u_x = -\frac{e\mathcal{E}}{m}t + U_0 \qquad x = -\frac{e\mathcal{E}}{2m}t^2 + U_0 t + X_0$$

◆ In any magnetic field, $\mathbf{f}$ is normal to $\mathbf{u}$ and no work is done.
In a uniform magnetic field, the path is circular with $r = mU_0/eB$.

◆ A cathode-ray tube consists of an electron gun producing a focused beam, a magnetic or electric deflection system, and a fluorescent screen for visual display.
A cathode-ray oscilloscope measures voltage and time, displays waveforms, and compares phase and frequency.

◆ Electron emission from a solid requires the addition of energy equal to the work function; this energy can be obtained in various ways.
The probability of an electron possessing energy eV_T varies as $e^{-eV_T/kT}$.

◆ In the vacuum diode, current is limited by the space charge and

$$I_b \cong KV_b^{3/2}$$

In a gas diode, cumulative ionization occurs, positive space charge is produced, and voltage drop is small.
In a phototube, electron emission is proportional to illumination.

◆ Multielectrode vacuum tubes possess basic amplifier characteristics.
In a triode, a small grid voltage modulates the space charge, controls the plate current, and develops an amplified load voltage.
Tetrodes and pentodes provide improved external characteristics.

◆ Graphical analysis of highly nonlinear tubes is based on load-line construction. Variations in input quantities about a properly selected quiescent point result in corresponding output variations of greater magnitude.

REVIEW QUESTIONS

1. What quantities are analogous in the equations of motion of a mass in a gravitational field and motion of an electron in an electric field?
2. Define the following terms: electron-volt, electron gun, vector cross-product, work function, space-charge limited, and virtual cathode.
3. Justify the statement that "no work is done on a charged particle by a magnetic field."
4. Describe the motion of an electron starting from rest in a region of parallel electric and magnetic fields.
5. Sketch the pattern expected on a C-R tube screen when the deflection voltages are $v_x = V \sin \omega t$ and $v_y = V \cos 2\omega t$.
6. What is the effect on a TV picture of varying the voltage on the control anode (Fig. 10.6)? On the accelerating anode?
7. Sketch a magnetic deflection system for a C-R tube.
8. Explain qualitatively the process of thermionic emission.
9. Explain the effect of space charge on diode current.
10. What happens in a diode operating normally if the cathode heating power is slowly reduced?
11. If $v(x) = V_m x^n$, is d^2v/dx^2 proportional to V_m?
12. Why is the voltage drop across a gas diode lower than across a similar vacuum diode for the same current?
13. Does the color of the light affect the curves of Fig. 10.18?
14. Sketch, from memory, the v-i characteristics of an ideal voltage amplifier; superimpose the characteristics of a triode.
15. Explain the effect of the grid on the electric field at the cathode.
16. What is the power amplification of a negative grid triode?
17. Describe how voltage amplification is achieved by a triode.
18. Compared to a triode, what is the advantage of a tetrode? Of a pentode?
19. Explain the function of control, screen, and suppressor grids.
20. Sketch, from memory, the characteristics of an ideal current amplifier; superimpose the characteristics of a pentode.
21. Explain the role of the load-line in graphical analysis of amplifiers.
22. In an amplifier, what is the function of V_{bb}, V_{cc}, R_L?
23. Define: quiescent point, grid swing, signal current, voltage gain, linear amplification.

EXERCISES

1. An electron is accelerated from rest in a region where $\mathcal{E} = 20$ V/m. Predict the velocity, kinetic energy gained, displacement, and potential energy lost after 2 ns. Compare the PE lost ($e\,\Delta V$) and the KE gained ($\frac{1}{2}mu^2$).
2. An electron is to be accelerated from rest to a velocity $u = 8 \times 10^6$ m/s after traveling a distance of 5 cm.
   (a) Specify the electric field required and estimate the time required.
   (b) Repeat part (a) for a hydrogen ion.
   (c) Express the energy gained by each particle in joules.
3. For the electron of Exercise 1, derive expressions for velocity as a function of time and as a function of distance.
4. An electron with an energy of 200 eV is projected at an angle of 45° into the region between two parallel plates carrying a voltage V and separated a distance $d = 5$ cm (see Fig. 10.29).
   (a) If $V = -200$ V, determine where the

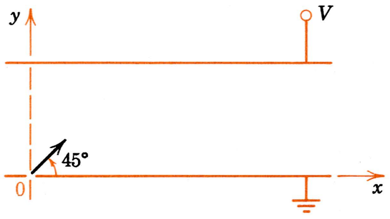


Figure 10.29

electron will strike.

(b) Determine the voltage V at which the electron will just graze the upper plate.

5. If the polarity of the upper plate in Fig. 10.29 is reversed so that $V = +200$ V, determine where the electron will strike.

6. Two electrons, e_1 traveling at velocity u and e_2 traveling at velocity $2u$, enter an intense magnetic field directed out of the paper (Fig. 10-30). Sketch the paths of the electrons.

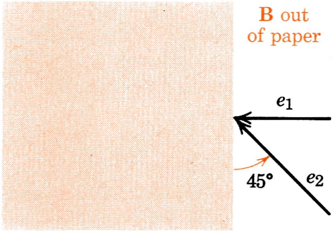


Figure 10.30

7. An electron with a velocity of 3000 km/s is injected at right angles to a uniform magnetic field where $B = 0.02$ T out of the paper.

(a) Sketch the path of the electron.

(b) Predict the radius of the path and the time required to traverse a semicircle.

8. In a *mass spectrograph* isotopes of charge e and unknown mass are accelerated through voltage V and injected into a transverse magnetic field B. Derive an expression for the mass m in terms of the radius r of the circle described by the particle.

9. In a cyclotron (see any physics book) hydrogen ions are accelerated by an electric field, curved around by a magnetic field, and accelerated again. If B is limited to 2 T, what cyclotron diameter is required for a velocity half that of light?

10. A C-R tube is operated with an accelerating potential of 5 kV. The vertical deflecting plates are 2 cm long and 1 cm apart; length L_S in Fig. 10.7 is 50 cm.

(a) Determine the deflection sensitivity.

(b) What is the maximum allowable deflection voltage?

(c) For a brighter picture, the acceleration potential is increased to 7.5 kV. What is the new deflection sensitivity?

11. How long is an electron in the deflecting region of the C-R tube of Exercise 10? If the deflecting voltage should not change more than 10% while deflection is taking place, approximately what frequency limit is placed on this deflection system?

12. In the CRT deflection system of Fig. 10.7, the accelerating voltage is 1000 V, the deflecting plates are 2.5 cm long and 1 cm apart, and the distance to the screen is 40 cm.

(a) Determine the velocity of the electrons striking the screen.

(b) Determine the deflecting voltage required for a deflection of 4 cm.

13. A CRO with a 10 × 10 cm display has vertical amplifier settings of 0.1, 0.2, and 0.5 V/cm and horizontal sweep settings of 1, 2, and 5 ms/cm. Assuming the sweep starts at $t = 0$, select appropriate settings and sketch the pattern observed when a voltage $v = 2 \sin 60\pi t$ V is applied to the vertical input.

14. Repeat Exercise 13 for an applied voltage $v = 0.5 \cos 1000t$ V.

15. On a CRO vertical and horizontal amplifiers are set at 1 V/cm. Sketch the pattern observed when the voltages applied to the vertical and horizontal inputs are

(a) $v_v = 5 \cos 400t$ and $v_h = 5 \cos 100t$ V.

(b) $v_v = 5 \cos 200t$ and
$v_h = 5 \cos (700t - \pi/2)$ V.

16. Design a circuit for displaying the v-i characteristic of a diode on a CRO. Assume that the horizontal and vertical amplifiers have a common internal ground.

17. The work function of copper is 4.1 eV and the melting point is 1356 K. Thorium has a work function of 3.5 eV and melts at 2120 K. Compare the factor of $e^{-W_W/kT}$ for these two metals at 90% of their respective melting points and decide which is more likely to be used as a cathode.

18. The Richardson-Dushman equation indicates that cathode emission current is proportional to $T^2 e^{-W_W/kT}$. For a special cathode coating with a work function of 1 eV, the emission current density at 1100 K is 200 mA/cm^2. Estimate the emission current densities at 800 and 1200 K.

19. The anode voltage in a diode is $+50$ V. As a result of space charge, a voltage minimum of -2 V exists at a distance of 1 mm from the plane cathode. Estimate the minimum velocity with which an electron can leave the cathode surface in order to arrive at the anode.

20. For a given vacuum diode it is known that $I_b = 5\text{ mA}$ when $V_b = 25\text{ V}$.
(a) *Stating* any necessary assumptions, predict the current at $V_b = 100\text{ V}$ and sketch I_b vs V_b to scale.
(b) The diode is connected in the circuit of Fig. 10.31. Predict diode current I_D.

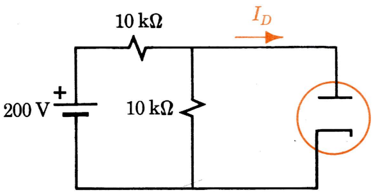


Figure 10.31

21. A vacuum diode can be considered to be a pair of parallel-plane electrodes separated by 0.6 cm. The velocity distribution across the interelectrode space is given in Fig. 10.32.

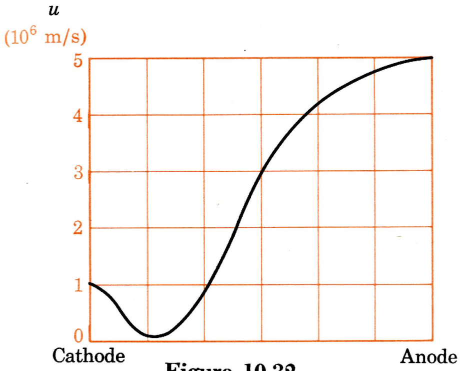


Figure 10.32

(a) What is the typical emission energy in electron volts?
(b) Assuming the cathode is at zero potential, what is the anode potential?
(c) Where is the virtual cathode with respect to the actual cathode?
(d) Sketch the potential distribution (clearly labeled) across the space.
(e) Sketch the distribution of charge density.

22. For a vacuum diode (or triode) to be useful at high frequencies, the transit time of the electrons from cathode to plate should be small compared to the period of the voltage waveform. Assuming a plate voltage of 50 V, a cathode-anode spacing of 3 mm, and assuming a linear potential distribution between plate and anode,
(a) Estimate the highest frequency for normal operation of the diode.
(b) If you had included the effect of space charge, would the transit time have been larger or smaller? Explain.

23. A diode has the characteristic of the $v_c = 0$ curve in Fig. 10.24b. Predict the current that will flow if the diode is connected in series with a 20-kΩ resistance and a 300-V battery.

24. In a practical diode, $I_b = 10\text{ mA}$ when $V_b = 25\text{ V}$.
(a) Derive an approximate expression for $I_b = f(V_b)$.
(b) Predict the anode current for $V_b = 100\text{ V}$.
(c) Predict the reading on a dc milliammeter connected in series with the diode, a resistance of 5 kΩ, and a voltage $v = 100\text{ V}$.

25. A mercury vapor diode (Fig. 10.17) is connected in series with a 1-kΩ resistance and a 30-V battery (negative terminal connected to the cathode of the diode). Sketch the circuit and predict the current in the resistance.

26. In an ideal vacuum tube with $\mu = 20$, $i_b = 10\text{ mA}$ for $v_c = 0$ and $v_b = 100\text{ V}$.
(a) Sketch the i_b vs v_b characteristics for $v_c = 0, -5\text{ V}$, etc.
(b) If $V_{bb} = 200\text{ V}$ and $R_L = 20\text{ k}\Omega$, plot the "dynamic transfer characteristic" consisting of i_b vs v_c for the range of nonpositive v_c for which plate current flows.

Note: Figures A1–A8 are located in the Appendix.

27. Consider the 6AU6-A pentode of Fig. A6 as an approximation to the idealized amplifier of Fig. 10.20. A load resistance of 20 kΩ is connected in the plate circuit and an average plate current of 8 mA is maintained. Estimate the voltage amplification possibilities of this transducer.

28. A 12AX7 triode (Fig. A5) is used in the circuit of Fig. 10.27a with $V_{cc} = 2.5$ V (polarity as shown), $V_{bb} = 400$ V, and $R_L = 125$ kΩ.

(a) Find i_b (graphically) for $v_i = 0$, $+2$ and -2 V.

(b) If v_i is a sinusoid varying between $+2$ and -2 V, what is the corresponding incremental variation in i_b?

(c) What is the corresponding incremental variation in v_L.

(d) Is the circuit performing amplification? What is the voltage gain?

29. A 6AQ5-A beam power tube (Fig. A8) is to be used as an elementary amplifier to provide an output signal of 300 V "peak to peak." The quiescent point is to be at $I_b = 50$ mA and $V_b = 200$ V.

(a) Draw a load line for $R_L = 4$ kΩ and determine the necessary supply voltages V_{bb} and V_{cc}.

(b) Estimate graphically the necessary grid signal $v_i(t)$.

(c) Calculate the voltage gain of the amplifier.

PROBLEMS

1. As Einstein pointed out, the actual mass m of a moving particle is $m = m_0/\sqrt{(1 - u^2/c^2}$ where m_0 is the rest mass, u is velocity, and c is the velocity of light. Since energy and mass are equivalent, the potential energy lost in falling through a voltage V must appear as an increase in mass $m - m_o$.

(a) Equate the potential energy lost to the equivalent energy gained and calculate the voltage required to accelerate an electron to 99% of the speed of light.

(b) Calculate the electron velocity for $V = 24$ kV, a typical value in a television picture tube, and determine the percentage of error in Eq. 10-11.

2. The magnetic deflecting *yoke* of a TV picture tube provides a field $B = 0.002$ T over an axial length $l = 2$ cm. The *gun* provides electrons with a velocity $U_Z = 10^8$ m/s. Determine:

(a) The radius of curvature of the electron path in the magnetic field.

(b) The approximate angle at which electrons leave the deflecting field.

(c) The distance from yoke to screen for a 3-cm *positive* deflection.

(d) A general expression for deflection Y in terms of acceleration potential V_a and compare with Eq. 10-20. (Assume small angular deflections where $Y/L = \tan\alpha \cong \alpha$.)

3. Under space-charge limited conditions, the potential distribution in a parallel-plane diode is $v/V_b = (x/d)^{4/3}$, where d is the electrode spacing. Show that an electron leaving the cathode with zero velocity under space-charge conditions takes 50% longer to get to the plate than in the absence of space charge.

4. A vacuum diode for which $I = 100$ mA at $V = 250$ V is temperature-limited to 250 mA.

(a) Sketch approximately the I-V characteristic for this diode.

(b) If a voltage $v = 250 + 20\cos\omega t$ V is applied across the diode, approximately what current $i(t)$ will flow?

(c) If a voltage $v = 250 + 350\cos\omega t$ V is applied across the diode, sketch, approximately to scale, the resulting current as a function of angle ωt.

5. A vacuum diode (Fig. 10.13a) consists of parallel plane electrodes separated a distance of 1 mm. Under space-charge limited conditions, a current of 50 mA flows when the plate voltage is 36 V.

(a) Predict the current for a plate voltage of 100 V.

(b) Predict the current at a plate voltage of 36 V in an otherwise identical diode with a cathode-plate spacing of 2 mm. (Explain your reasoning.)

6. For the phototube characteristics of Fig. 10.18, the scales are 2 μA/cm for I_b and 50 V/cm for V_b. Design a circuit so that with a battery voltage $V_{bb} = 250$ V and a steady illumination of 0.2 lumen, a voltage of approximately 100 V will appear across load resistor R_L. Determine the ac component of the output voltage for an illumination signal of $0.2 + 0.1\sin\omega t$ lumens.

Chapter 11

Semiconductor Diodes

Conduction in Solids
Doped Semiconductors
Junction Diodes
Special Diodes

The multielectrode vacuum tube, as exemplified by the pentode and the cathode-ray tube, is a versatile means for precisely controlling electron flow. Such tubes have been widely used in a host of applications. However, the high power necessary for thermionic emission and the large surfaces required for practical operation are important disadvantages. In contrast, semiconductor electronic devices require no cathode power and their dimensions are very small. As a result, the vacuum tube has been displaced from many applications in which it was formerly preeminent, and the emphasis in the rest of this book will be on semiconductor diodes and transistors in discrete and integrated forms.

First we consider conduction in solids where electron motion is influenced by the fixed ions of a conductor or the doping atoms added to a semiconductor. Using our knowledge of conduction mechanisms, we then examine the operation of semiconductor diodes and develop a quantitative understanding of diode junction behavior. We shall use our knowledge of junctions in considering diode applications in Chapter 12 and in analyzing the operation of transistors in Chapter 13.

Conduction in Solids

Conduction occurs in a vacuum if free electrons are available to carry charge under the action of an applied field. In an ionized gas, positively charged ions as well as electrons contribute to the conduction process. In a liquid, the charge carriers are positive and negative ions moving under the influence of an applied field. Solids vary widely in the type and number of charge carriers available, and in the ease with which the carriers move under the action of applied fields. In electrical engineering we are interested in *insulators*, which have practically no available charge carriers; *conductors,* which have large numbers of mobile charge carriers; and *semiconductors,* which have conductivities intermediate between those of insulators and conductors.

Metallic Conductors

In a typical metal such as copper or silver, the atoms are arranged in a systematic array to form a *crystal.* The atoms are in such close proximity that the outer, loosely bound electrons are attracted to numerous neighboring nuclei and, therefore, are not closely associated with any one nucleus (Fig. 11.1). These *conduction electrons* are visualized as being free to wander through the crystal structure or *lattice.* At absolute zero temperature, the conduction electrons encounter no opposition to motion and the resistance is zero.

At ordinary temperatures, the electron-deficient atoms called *ions* possess kinetic energy in the form of vibration about neutral positions in the lattice; this vibrational energy is measured by temperature. There is a continual interchange of energy between the vibrating ions and the free electrons in the form of elastic and inelastic collisions. The resulting electron motion is random, there is no net motion, and the net current is zero (Fig. 11.2a).

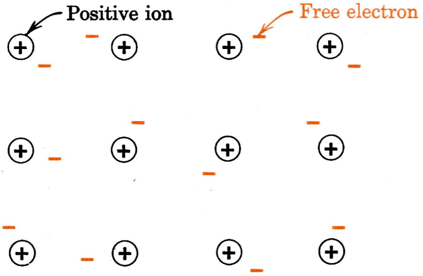
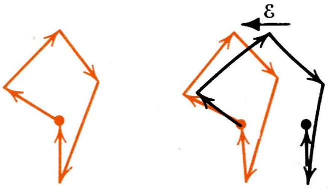


Figure 11.1 A simplified two-dimensional representation of a metallic crystal.



(a) (b)

Figure 11.2 The motion of an electron in a metallic crystal.

If a uniform electric field of intensity $\mathcal{E}$(V/m) is applied, the electrons are accelerated; superimposed on the rapid random motion, there is a small component of velocity in the direction of $-\mathcal{E}$. At each inelastic collision with an ion, the electron loses its kinetic energy; it then accelerates again, gains a component of velocity in the $-\mathcal{E}$ direction, and loses its energy at the next inelastic collision (Fig. 11.2b). The time between collisions is determined by the random velocity and the length of the mean free path. On the average, the electrons gain a directed *drift velocity u* that is directly proportional to $\mathcal{E}$ (since the acceleration is constant for an average increment of time), and

$$u = \mu\mathcal{E} \tag{11-1}$$

where μ (mu) is the *mobility* in meters per second/volts per meter or m^2/V-s.

The resulting flow of electrons carrying charge $-e$ at drift velocity u constitutes a current. If there are n free electrons per cubic meter, the current density J (A/m^2) is

$$J = neu = ne\mu\mathcal{E} = \sigma\mathcal{E} \tag{11-2}$$

where

$$\sigma = ne\mu$$

is the *conductivity* of the material in mhos per meter. The reciprocal of σ (sigma) is the *resistivity* ρ (rho) in ohm-meters. Equation 11-2 is another formulation of Ohm's law; it says that the current density is directly proportional to the voltage gradient. For a conductor of area A (m^2) and length l (m),

$$I = JA = \sigma A\mathcal{E} = \sigma A\frac{V}{l} = \sigma\frac{A}{l}V \tag{11-3}$$

where $\sigma A/l$ is the conductance G in mhos. Alternatively,

$$R = \frac{V}{I} = \frac{1}{G} = \frac{l}{\sigma A} = \rho\frac{l}{A} \tag{11-4}$$

where R is the resistance in ohms.

Example 1

The resistivity of copper at 20°C is $1.73 \times 10^{-8}\ \Omega$-m. Find the average drift velocity in a copper conductor with a cross-sectional area of 10^{-6} m^2 carrying a current of 4 A.

For copper with an atomic weight of 63.6 and a density of 8.9 g/cm^3, by Avogadro's law the number of atoms per meter^3 is

$$n_A = \frac{6.022 \times 10^{23}\ \text{atoms/g-atom} \times 8.9 \times 10^6\ \text{g/m}^3}{63.6\ \text{g/g-atom}}$$

$$= 8.43 \times 10^{28}$$

Assuming that there is one free electron per atom, $n = n_A$. By Eq.11-2,

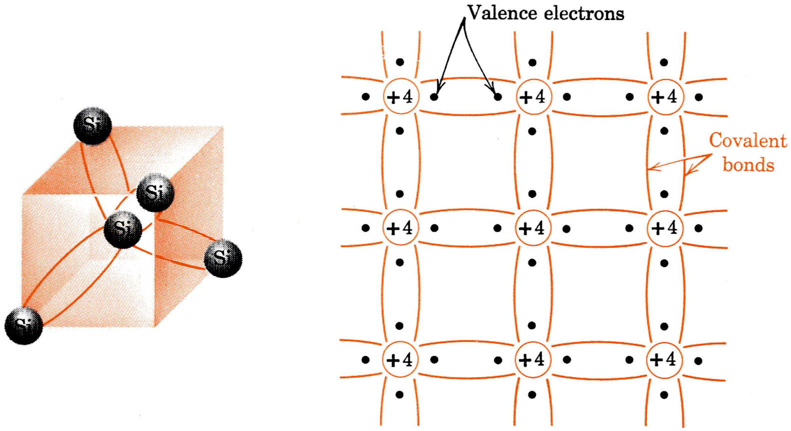
$$u = \frac{J}{ne} = \frac{I/A}{ne} = \frac{4 \times 10^6}{8.43 \times 10^{28} \times 1.6 \times 10^{-19}}$$

$$\cong 3 \times 10^{-4}\ \text{m/s}$$

The average drift velocity in a good conductor is very slow compared to the random thermal electron velocities of the order of 10^5 m/s at room temperatures. As temperature increases, random thermal motion increases, the time between energy-robbing collisions decreases, mobility decreases, and therefore conductivity decreases. It is characteristic of metallic conductors that resistance increases with temperature.

Semiconductors

The two semiconductors of greatest importance in electronics are silicon and germanium. These elements are located in the fourth column of the periodic table and have four valence electrons. The crystal structure of silicon or germanium follows a tetrahedral pattern with each atom sharing one valence electron with each of four neighboring atoms. The *covalent bonds* are shown in the two-dimensional representation of Fig. 11.3. At temperatures close to absolute zero, the electrons in the outer shell are tightly bound, there are no free carriers, and silicon is an insulator.



(a) Tetrahedral pattern **(b) Two-dimensional representation**

Figure 11.3 Arrangement of atoms in a silicon crystal.

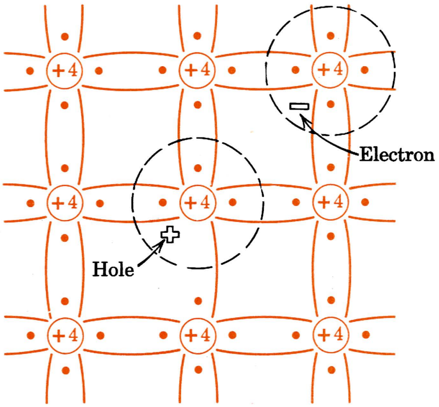


Figure 11.4 Silicon crystal with one covalent bond broken.

The energy required to break a covalent bond is about 1.1 eV for silicon and about 0.7 eV for germanium. At room temperature (300 K), a few electrons have this amount of thermal energy and are excited into the conduction band (see p. 267), becoming free electrons. When a covalent bond is broken (Fig. 11.4), a vacancy or *hole* is left. The region in which this vacancy exists has a net positive charge; the region in which the freed electron exists has a net negative charge. In such semiconductors *both electrons and holes contribute to electrical conduction.* If a valence electron from another covalent bond fills the hole (without ever gaining sufficient energy to become "free"), the vacancy appears in a new place and the effect is as if a positive charge (of magnitude e) has moved to the new location.†

Consideration of hole behavior as the motion of a positively charged particle of definite mass and mobility is consistent with the quantum mechanical properties of valence electrons. Attempting to describe this behavior in terms of classical physics leads to results that are contrary to the observed facts. For our purposes we shall consider that:

> *Conduction in semiconductors is due to two separate and independent particles carrying opposite charges and drifting in opposite directions under the influence of an applied electric field.*

With two charge-carrying particles, the expression for current density is

$$J = (n\mu_n + p\mu_p)e\mathcal{E} = \sigma\mathcal{E} \tag{11-5}$$

where n and p are the concentrations of electrons and holes (number/m^3) and μ_n and μ_p are the corresponding mobilities. Therefore, the conductivity is

$$\sigma = (n\mu_n + p\mu_p)e \tag{11-6}$$

Some of the properties of silicon and germanium are shown in Table 11-1.

Table 11-1 Properties of Silicon and Germanium at 300 K†

	Silicon	Germanium
Energy gap (eV)	1.1	0.67
Electron mobility μ_n (m^2/V-s)	0.135	0.39
Hole mobility μ_p (m^2/V-s)	0.048	0.19
Intrinsic carrier density n_i (/m^3)	1.5×10^{16}	2.4×10^{19}
Intrinsic resistivity ρ_i (Ω-m)	2300	0.46
Density (g/m^3)	2.33×10^6	5.32×10^6

†Quoted by Esther M. Conwell in "Properties of Silicon and Germanium II," *Proc. IRE,* June 1958, p. 1281.

Since holes and electrons are created simultaneously, in a pure semiconductor the number of holes is just equal to the number of conduction electrons, or

---

† There is an analogy in our economic system. In payment for a sweater, you may give a retailer $20 in cash or accept a bill from him for $20. You feel the same with $20 *subtracted* from your wallet or with $20 *added* to your debts. If you send the bill to your father for payment, the "vacancy appears in a new place" and the effect is as if your father had sent you $20. The sum of cash to you and bills to your father constitutes the net financial "current" from your father to you.

$$n = p = n_i \tag{11-7}$$

where n_i is the *intrinsic* concentration. New electron-hole pairs are generated continuously and, under equilibrium conditions, they are removed by recombination at the same rate. At ordinary operating temperatures, only a very small fraction of the valence electrons are in the conduction state at any instant.

Example 2

Estimate the relative concentration of silicon atoms and electron-hole pairs at room temperature, and predict the intrinsic resistivity.

By Avogadro's law, the concentration of atoms is

$$n_A = \frac{6.022 \times 10^{23}\,\text{atoms/g-atom} \times 2.33 \times 10^6\,\text{g/m}^3}{28.09\,\text{g/g-atom}}$$
$$\cong 5 \times 10^{28}\,\text{atoms/m}^3$$

From Table 11-1, the intrinsic concentration at 300 K is 1.5×10^{16} electron-hole pairs per cubic meter; therefore, there are

$$\frac{n_A}{n_i} \cong \frac{5 \times 10^{28}}{1.5 \times 10^{16}} \cong 3.3 \times 10^{12} \text{ silicon atoms per electron-hole pair}$$

Since in the pure semiconductor $n = p = n_i$, by Eq. 11-6 the intrinsic conductivity is

$$\begin{aligned} \sigma_i &= (n\mu_n + p\mu_p)e = (\mu_n + \mu_p)n_i e \\ &= (0.135 + 0.048) \times 1.5 \times 10^{16} \times 1.6 \times 10^{-19} \\ &= 4.4 \times 10^{-4}\,\mho/\text{m} \end{aligned}$$

and the intrinsic resistivity is

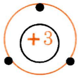
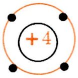
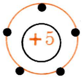
$$\rho_i = \frac{1}{\sigma_i} = \frac{1}{4.4 \times 10^{-4}} \cong 2280\ \Omega\text{-m}$$

The result of Example 2 confirms the fact that intrinsic silicon is a poor conductor as compared to copper ($\sigma = 5.8 \times 10^7\ \mho/\text{m}$); semiconductors are valuable, not for their conductivity, but for two unusual properties. First, the concentration of free carriers, and consequently the conductivity, increases exponentially with temperature (approximately 5% per degree centigrade at ordinary temperatures). The *thermistor*, a temperature-sensitive device that utilizes this property, is used in instrumentation and control, and also may be used to compensate for the *decrease* in conductivity with temperature in the metallic parts of a circuit. Second, the conductivity of a semiconductor can be increased greatly, and to a precisely controlled extent, by adding small amounts of impurities in the process called *doping*. Since there are two types of mobile charge carriers, of opposite sign, extraordinary distributions of charge carriers can be created. The semiconductor diode and the transistor utilize this property.

Doped Semiconductors

The startling effect of doping can be illustrated by adding to silicon or germanium an impurity element from the adjacent third or fifth column of the periodic table (see Table 11-2). Assume that a small amount of a *pentavalent* element (antimony, phosphorus, or arsenic) is added to otherwise pure silicon. Such a doping element has five valence electrons and the effective ionic charge is $+5e$. When a pentavalent atom replaces a silicon atom in the crystal lattice (Fig. 11.5), only four of the valence electrons are used to complete the covalent bonds; the remaining electron, with the addition of a small amount of thermal energy, becomes a free electron available for conduction. The resulting material is called an *n-type* semiconductor because of the presence of negative charge carriers in an electrically neutral crystal. The net positive charges left behind are immobile and cannot contribute to current.

Table 11-2 Semiconductor Elements in the Periodic Table (with atomic number and atomic weight)

III 	IV 	V 
5 B BORON 10.82	6 C CARBON 12.01	7 N NITROGEN 14.008
13 Al ALUMINUM 26.97	14 Si SILICON 28.09	15 P PHOSPHORUS 31.02
31 Ga GALLIUM 69.72	32 Ge GERMANIUM 72.60	33 As ARSENIC 74.91
49 In INDIUM 114.8	50 Sn TIN 118.7	51 Sb ANTIMONY 121.8

If a small amount of a *trivalent* element (aluminum, boron, gallium, or indium) is added to otherwise pure silicon, a *p-type* semiconductor is obtained. When

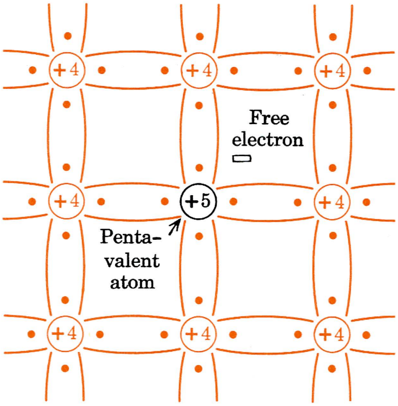


Figure 11.5 Effect of *n*-type doping.

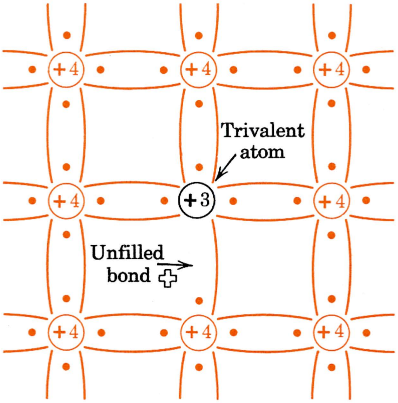


Figure 11.6 Effect of *p*-type doping.

a trivalent atom replaces a silicon atom in the crystal lattice (Fig. 11.6), only three valence electrons are available to complete covalent bonds. If the remaining unfilled covalent bond is filled by a valence electron from a neighboring atom, a mobile hole is created and there is the possibility of current conduction by the motion of positive charges. Used in this way, a trivalent atom is called an *acceptor* atom because it accepts an electron. (The net negative charge created is immobile.) In the same sense, a pentavalent atom is called a *donor* atom.

Conduction

By adding donor or acceptor atoms in small amounts, the conductivity of a semiconductor can be increased enormously. To predict the magnitude of this effect we must understand quantitatively the process of generation and recombination. The *thermal generation rate g* (electron-hole pairs/s-m^3) depends on the properties of the material and is a function of temperature. The energy required for a valence electron to become a free electron, or the energy needed in the "generation" of a conduction electron, can be expressed in electron volts as eV_g. The average energy corresponding to temperature T can be expressed as kT. Statistical analysis shows that the probability of a valence electron receiving sufficient energy to become free is proportional to $e^{-eV_g/kT}$; the rate of thermal generation (and also the rate of thermionic emission; see p. 268) involves this factor and, therefore, is highly temperature dependent.

In a semiconductor the mobile electrons and holes tend to recombine and disappear. If there are few electron-hole pairs in existence, the *rate of recombination* is low; if there are many, the rate is high. If, as in *n*-type material, there are few holes but many electrons, the rate of recombination is high due to the large number of electrons. In general,

$$R = rnp \tag{11-8}$$

where R = recombination rate (electron-hole pairs/s-m^3) and
r = a proportionality constant for the material.

This version of the *mass-action law* says that the rate of recombination is dependent on the numbers of reacting elements present.†

Under equilibrium conditions, the rate of generation is just equal to the rate of recombination or

$$g = R = rnp \tag{11-9}$$

In the pure crystal the intrinsic concentrations of electrons and holes are equal, or $n_i = p_i$ and

$$g = R = rn_ip_i = rn_i^2 \tag{11-10}$$

Even in a doped crystal, the great bulk of the atoms are still silicon (or germanium) and the thermal generation rate is unchanged from the intrinsic value. Therefore, Eq. 11-10 is a general relation at a given temperature and

$$np = n_i^2 \tag{11-11}$$

in doped semiconductors as well.

A second fundamental relation is based on the fact that the total crystal must be electrically neutral. There are two types of charge to consider: *immobile* ions and *mobile* carriers. Practically all donor or acceptor atoms are ionized at ordinary temperatures leaving immobile positive and negative ions in concentrations N_d and N_a. In addition there are mobile positive and negative carriers in concentrations p and n. For electrical neutrality in any doped crystal,

$$p + N_d = n + N_a \tag{11-12}$$

In n-type material created by adding donor impurity to intrinsic semiconductor, $N_a = 0$ and

$$n_n = N_d + p_n \cong N_d \tag{11-13a}$$

in the practical case where $N_d \gg p_n$. (See Example 3.) The concentration of holes in n-type material is, by Eq. 11-11,

$$p_n = \frac{n_i^2}{n_n} \cong \frac{n_i^2}{N_d} \tag{11-13b}$$

The corresponding equations for p-type material are:

$$p_p \cong N_a \quad \text{and} \quad n_p \cong \frac{n_i^2}{N_a} \tag{11-14}$$

These relations are used in predicting the effect of doping in Example 3.

---

† As a crude analogy, consider a dance in which couples dance together for awhile, then separate and wander among the other couples until they find a new partner. The rate of recombination depends on the number of unattached partners available.

Example 3

Predict the effect on the electrical properties of a silicon crystal at room temperature if every 10-millionth silicon atom is replaced by an atom of indium.

Since there are 5×10^{28} silicon atoms/m^3, the necessary concentration of acceptor atoms is

$$N_a = 5 \times 10^{28} \times 10^{-7} = 5 \times 10^{21} \cong p_p$$

The corresponding intrinsic concentration is $p_i = n_i = 1.5 \times 10^{16}/\text{m}^3$. Therefore, there are $(5 \times 10^{21}) \div (1.5 \times 10^{16}) = 3.3 \times 10^5$ as many holes. The new electron concentration is

$$n_p = \frac{n_i^2}{p_p} \cong \frac{n_i^2}{N_a} = \frac{2.25 \times 10^{32}}{5 \times 10^{21}} = 4.5 \times 10^{10}/\text{m}^3$$

Therefore, the electron concentration has been reduced by the factor $(1.5 \times 10^{16}) \div (4.5 \times 10^{10}) = 3.3 \times 10^5$ and Eq. 11-13a is justified.

The new conductivity is

$$\begin{aligned}\sigma &= (n\mu_n + p\mu_p)e \cong N_a(\mu_p)e \\ &= 5 \times 10^{21} \times 0.048 \times 1.6 \times 10^{-19} = 38.4\ \mho/\text{m}\end{aligned}$$

as compared to $0.00044\ \mho/\text{m}$ in intrinsic silicon. In other words, the resistivity of the doped silicon is

$$\rho = 1/\sigma = 1/38.4 = 0.026\ \Omega\text{-m}$$

as compared to the intrinsic value $\rho_i = 2300\ \Omega\text{-m}$ in Table 11.1.

We conclude from Example 3 that a small concentration of impurity greatly modifies the electrical properties of a semiconductor. Even more startling effects occur in regions where the impurity concentration is highly nonuniform, as it is at a junction of *n*-type and *p*-type materials.

Diffusion

If the doping concentration is nonuniform, the concentration of charged particles is also nonuniform, and it is possible to have charge motion by the mechanism called *diffusion*. In an analogous situation, if a bottle of perfume is opened at one end of an apparently still room, in a short time some of the perfume will have reached the other end. For mass transfer by diffusion, two factors must be present: There must be random motion of the particles and there must be a concentration gradient. The thermal energy of apparently still air provides the first, and opening a bottle of concentrated perfume provides the second. Note that no force is required for this type of transport; the net transfer is due to a statistical redistribution.

If a nonuniform concentration of randomly moving electrons or holes exists, diffusion will occur. If, for example, the concentration of electrons on one side of an imaginary plane is greater than that on the other, and if the electrons are in random motion, after a finite time there will be a net motion of electrons from the more dense side. This net motion of charge which constitutes a *diffusion current* is proportional to the concentration gradient dn/dx. The diffusion current density due to electrons is given by

$$J_n = eD_n \frac{dn}{dx} \tag{11-15}$$

where D_n = the diffusion constant for electrons (m^2/s). If dn/dx is positive, the resulting motion of negatively charged electrons in the $-x$ direction constitutes a positive current in the $+x$ direction. The corresponding relation for holes is

$$J_p = -eD_p \frac{dp}{dx} \tag{11-16}$$

Note that in each case the charged particles move away from the region of greatest concentration, but the motion is not due to any force of repulsion. Like mobility, diffusion is a statistical phenomenon and it is not surprising to find that they are related by the Einstein equation

$$\frac{\mu_n}{D_n} = \frac{\mu_p}{D_p} = \frac{e}{kT} \tag{11-17}$$

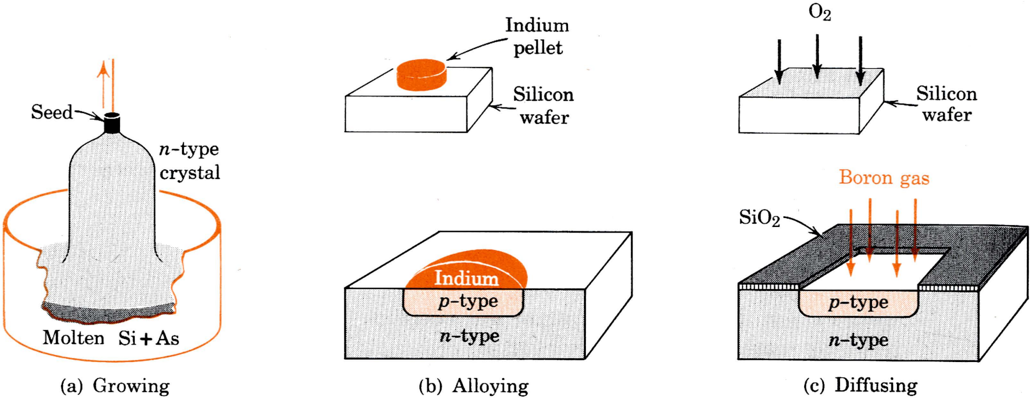
or that the factor e/kT, which appeared in thermionic emission and thermal generation, shows up again here.

Junction Diodes

Semiconductors, pure or doped, p-type or n-type, are bilateral; current flows in either direction with equal facility. If, however, a p-type region exists in close proximity to an n-type region, there is a carrier density gradient that is unilateral; current flows easily in one direction only. The resulting device, a *semiconductor diode,* exhibits a very useful control property.

pn Junction Formation

The fabrication of semiconductor diodes is a developing technological art. For satisfactory results, the entire diode must have a common crystalline structure. The pure crystals necessary may be "grown" by touching a "seed" crystal to the surface of molten silicon and then slowly raising it (Fig. 11.7a). The silicon crystallizes onto the seed and "grows" in the same crystalline orientation as the seed. If the melt contains a donor element such as arsenic, n-type semiconductor is obtained. If at a certain point in the growing process an excess of acceptor element is added to the melt, the subsequent crystal will be p-type. The metallurgical boundary between the two materials is called a *grown junction.*



(a) Growing (b) Alloying (c) Diffusing

Figure 11.7 Methods of forming *pn* junctions.

If a dot of trivalent indium is placed on an n-type silicon wafer (sliced from a grown crystal) and heated to the proper temperature, a small drop of molten mixture will form. Upon cooling, the alloy solidifies into p-type silicon that follows the crystalline pattern of the host crystal (Fig. 11.7b). The resulting device is an *alloy-junction* diode. By careful control of the alloying procedure it is possible to obtain diodes in which the transition from p-type to n-type material occurs in a distance corresponding to a few atomic layers.

Precisely controlled geometries and improved device properties are possible with the *planar* technology illustrated in Fig. 11.7c. First a thin layer of an inert oxide (such as glassy silicon dioxide) is formed on the surface of the n-type semiconductor wafer by heating in the presence of oxygen. Then the oxide layer is removed from a selected area by etching through a mask. When the heated semiconductor is exposed to boron gas, for example, acceptor atoms penetrate the selected area of the crystalline structure by solid-state diffusion and a precisely defined *pn* junction is formed. (The unetched oxide layer is impervious.) In another technique, the junction is formed by deposition of the doped material from a vapor. The atoms formed by this *epitaxial* growth (grown "upon, in the same form") follow the orientation of the underlying *substrate* crystal, but may possess entirely different electrical characteristics.

pn Junction Behavior

The external behavior of a typical *pn* junction diode is shown in Fig. 11.8. To use a diode effectively, we must understand (at least qualitatively) the internal mechanism that results in these remarkable properties. An understanding of diode behavior is directly transferable to the analysis of more complicated devices, such as the transistor, so a detailed study at this point is justified. We assume that the diode consists of a single crystal with an abrupt change from p- to n-type at the junction in the y-z plane; only variations with respect to x are significant.

The distribution of charges near the junction in such a diode is represented in Fig. 11.9. The circles represent immobile ions with their net charges; the separate + and − signs represent mobile holes and electrons. We imagine that a junction is created instantaneously; because of the density gradient at the junction, holes diffuse from the *p*-type material to the right across the junction and recombine with some of the many free electrons in the *n*-type material. Similarly, electrons diffuse from the *n*-type material to the left across the junction and recombine. The diffusion of holes leaves *uncovered* bound negative charges on the left of the junction, and the diffusion of electrons leaves *uncovered* bound positive charges on the right.

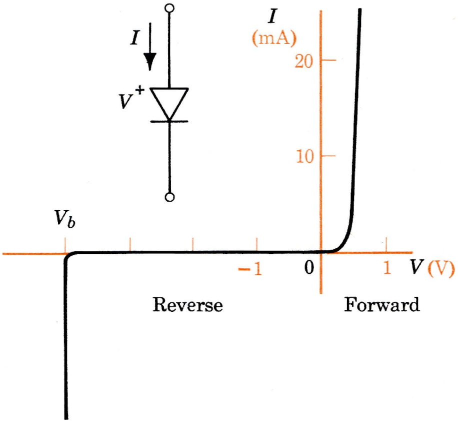


Figure 11.8 Junction diode symbol and *I-V* characteristic for germanium.

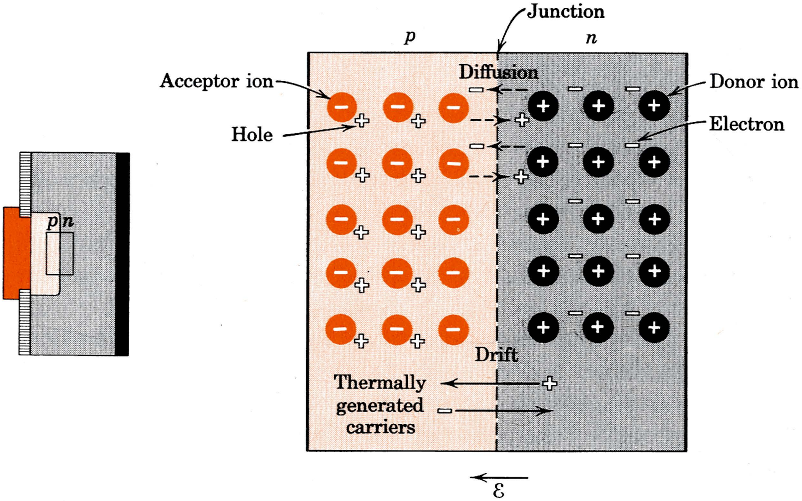


Figure 11.9 Schematic representation of charges and conduction mechanisms in a section of a *pn* diode.

To avoid repeating every statement, we call holes in the *p*-region and electrons in the *n*-region *majority carriers*. Thus the result of the diffusion of majority carriers is the uncovering of bound charge in the *depletion* or *transition region*

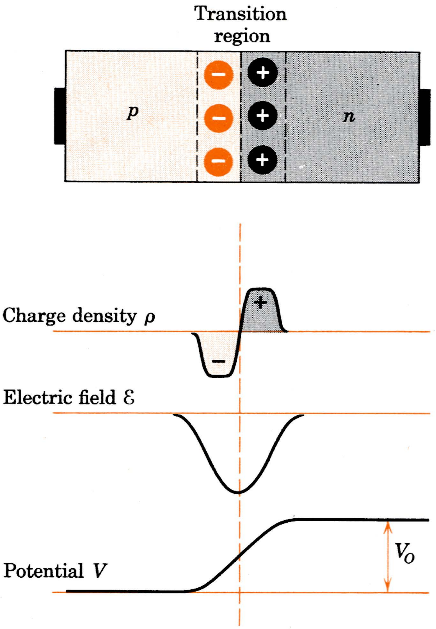


Figure 11.10 An open-circuited *pn* diode.

and the creation of an internal electric field. Electrons in the *p*-region and holes in the *n*-region are called *minority carriers*. Under the action of the electric field, minority carriers drift across the junction. With two types of carriers (reduced from four by the use of the terms "majority" and "minority") and two different conduction mechanisms (drift and diffusion) we expect the explanation of Fig. 11.8 to be complicated.

Open Circuit. The simpler representation of the diode in Fig. 11.10 shows only the uncovered charge in the transition region; the remainder of the open-circuited diode is electrically neutral. With the distribution of charge density shown, there is an electric field (see Eq. 10-27)

$$\mathcal{E} = \frac{1}{\epsilon}\int \rho \, dx \qquad (11\text{-}18)$$

Since $\mathcal{E} = -dv/dx$ (Eq. 10-25), the potential distribution across the junction is obtained by integrating the field. The potential barrier or "potential hill" of height V_O acts to oppose diffusion of majority carriers (holes to the right and electrons to the left) and to encourage drift of minority carriers across the junction.

Under open-circuit conditions the net current is zero; the tendency of majority carriers to diffuse as a result of the density gradient is just balanced by the tendency of minority carriers to drift across the junction as a result of the electric field. The height of the potential barrier and the magnitude of V_O is determined by the necessity for equilibrium; for no net current, the diffusion component must be just equal to the drift component. The equilibrium value V_O is also called the *contact potential* for the junction. This potential, typically a few tenths of a volt, cannot be used to cause external current flow. Connecting an external conductor across the terminals creates two new contact potentials that just cancel that at the junction.†

---

† An analogy may be helpful. In a vertical column of gas, molecules drift downward under the action of gravity, but diffuse upward because of their random thermal energy. A pressure difference exists between top and bottom, but there is no net flow. The pressure difference cannot be used to cause any external flow.

Forward Bias. The diffusion component of current is very sensitive to the barrier height; only those majority carriers having kinetic energies in excess of eV_0 diffuse across the junction. (The probability of a carrier possessing sufficient energy can be expressed in terms of $e^{-eV_0/kT}$.) If equilibrium is upset by a decrease in the junction potential to $V_0 - V$ (Fig. 11.11), the probability of majority carriers possessing sufficient energy to pass the potential barrier is greatly increased. The result is a net flow of carriers. Holes from *p*-type material cross the junction and are injected into *n*-type material where they represent minority carriers; electrons move in the reverse direction.

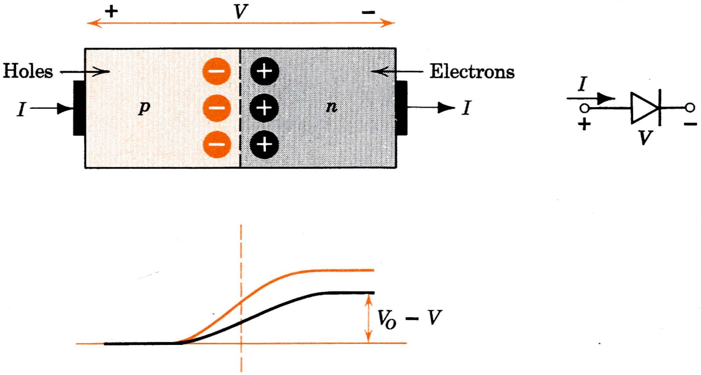


Figure 11.11 A forward-biased *pn* diode.

In other words, a reduction in the potential barrier encourages carriers to flow from regions where they are in the majority to regions where they are in the minority. If the junction potential is reduced by the application of an external *biasing* voltage *V*, as shown in Fig. 11.11, a net current *I* flows. (The small voltage drop across the diode body at moderate currents is neglected in this discussion.) Under forward bias, the current is sustained by the continuous generation of majority carriers and recombination of the injected minority carriers.

Reverse Bias. If the junction potential is raised to $V_0 + V$ by the application of a reverse bias (Fig. 11.12), the probability of majority carriers possessing sufficient energy to pass the potential barrier is greatly decreased. The net diffusion of majority carriers across the junction, and therefore the minority carrier injection current, is reduced to practically zero by a reverse bias of a tenth of a volt or so.

A very small reverse current does flow. The minority carriers thermally generated in the material adjacent to the transition region drift in the direction of the electric field. Since all minority carriers appearing at the edge of the transition region are forced across by the field (holes "roll down" a potential

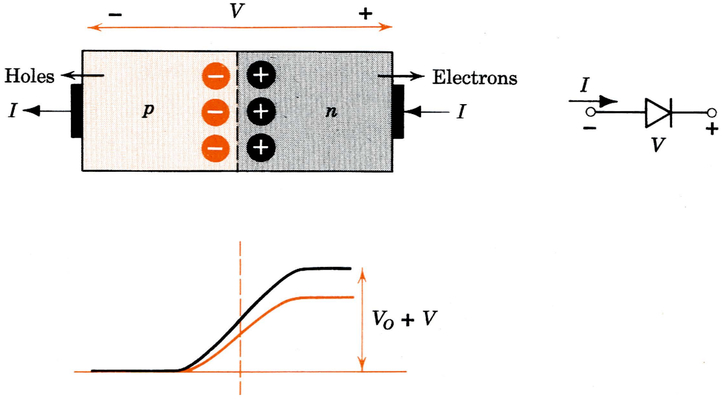


Figure 11.12 A reverse-biased *pn* diode.

hill; electrons "roll up"), this reverse *saturation* current I_s depends only on the rate of thermal generation and is independent of the barrier height. (This current reaches its maximum value or "saturates" at a low value of reverse bias.)

At ordinary temperatures the saturation current is very small, measured in μA (germanium) or nA (silicon). Ideally the reverse current in a diode should be zero; a low value of I_s is one of the principal advantages of silicon over germanium. Practically I_s is an important parameter that is sensitive to temperature. It is directly proportional to the thermal generation rate $g = rn_i^2$ which varies exponentially with temperature. As a rough approximation near room temperature, for silicon or germanium diodes, I_s doubles with each increase of 10°C.

General Behavior

Calculation of the current flow in a *pn* junction diode is usually based on a detailed analysis of the minority-carrier density gradients at the edges of the transition region.† For our purposes here, we can assume that the diode current consists of an injection component I_i sensitive to junction potential and a reverse saturation component I_s independent of junction potential. As we expect from our previous experience with thermionic emission and thermal generation, the injection process is related to the statistical probability of electrons and holes having sufficient energy. Under a forward bias voltage V, this probability increases; the critical factor becomes

$$e^{-e(V_O-V)/kT} = e^{-eV_O/kT} \cdot e^{eV/kT} = A_1 e^{eV/kT} \tag{11-19}$$

At a given temperature, the injection current I_i is of the form $A_1 A_2 e^{eV/kT} =$

---

† See pp. 32–42 of P. E. Gray, et al., *Physical Electronics and Circuit Models of Transistors,* John Wiley & Sons, New York, 1964.

$Ae^{eV/kT}$, and the saturation current I_s is negative and nearly constant. The total current is

$$I = I_i - I_s = Ae^{eV/kT} - I_s \tag{11-20}$$

Under open-circuit conditions where $V = 0$,

$$I = 0 = Ae^0 - I_s \quad \therefore A = I_s$$

In general, then, the junction diode current is

$$I = I_s(e^{eV/kT} - 1) \tag{11-21}$$

At room temperature (say, $T = 20°\text{C} = 293\text{ K}$), $e/kT \cong 40\text{ V}^{-1}$ and a convenient form of the equation is

$$I = I_s(e^{40V} - 1) \tag{11-22}$$

Example 4

The current of a germanium diode at room temperature is 100 μA at a voltage of -1 V. Predict the magnitude of the current for voltages of -0.2 V and $+0.2$ V at room temperature.

The current at -1 V can be assumed to be the reverse saturation current or $I_s = 100\ \mu\text{A}$.
At $V = -0.2$ V, $e^{40V} = e^{-8} \cong 0$.
Therefore, $I \cong -I_s = -100\ \mu\text{A}$.
At $V = +0.2$ V, $e^{40V} = e^{8} \cong 3000$.
Therefore, $I = 100(3000 - 1) \times 10^{-6} \cong +300$ mA.

Repeat the prediction for operation at 20°C above room temperature.

At 20°C above room temperature, I_s will double twice and be approximately four times as great. The exponent will be $8 \times 293/(293 + 20) = 7.49$. The two values of current will be $-I_s = -400\ \mu\text{A}$ and $I = +715$ mA, respectively.

We can draw some general conclusions from Example 4. Since $e^{40(-0.1)} = e^{-4} = 0.02$, for a reverse bias of more than 0.1 V or so, $I \cong -I_s$. Since $e^{40(0.1)} = e^{+4} = 55$, for a forward bias of more than 0.1 V or so, $I \cong I_s e^{40V}$.

Equation 11-21 holds quite well for germanium diodes, but only approximately for silicon diodes. At normal currents, silicon diodes follow the relation

$$I = I_s(e^{eV/\eta kT} - 1) \tag{11-23}$$

where factor η (eta) is approximately 2. For silicon diodes at high currents (and for germanium diodes in general), $\eta \cong 1$. (See Example 5 on next page.)

Special Purpose Diodes

The *pn* junction diode described by Eq. 11-21 is widely used in rectifying and waveshaping functions as discussed in the next chapter. Other diodes,† similar

---

† Millman and Halkias, *Integrated Electronics*, McGraw-Hill Book Co., New York, 1972, has an excellent discussion.

Example 5

The voltage at which significant diode current flows is called the *threshold voltage*. Assuming that the criterion is 1% of rated current, estimate the threshold voltage at room temperature for two diodes rated at 1 A: a germanium diode with $I_s = 3\,\mu\text{A}$ and a silicon diode with $I_s = 60\text{ nA}$.

By Eq. 11.22, for the germanium unit.

$$V \cong \frac{1}{40} \ln \frac{I}{I_s} = \frac{1}{40} \ln \frac{10^{-2}}{3 \times 10^{-6}} \cong 0.2\text{ V}$$

By Eq. 11.23, for the silicon unit with $\eta = 2$,

$$V \cong \frac{2}{40} \ln \frac{I}{I_s} = \frac{1}{20} \ln \frac{10^{-2}}{6 \times 10^{-8}} \cong 0.6\text{ V}$$

These are typical values for practical diodes.

in basic structure but presenting quite different external characteristics, can be explained in terms of our knowledge of junctions.

Breakdown Diodes

Along with the change in junction potential due to forward and reverse biasing, there is a significant change in the width of the transition or depletion region. For a given junction potential, a definite number of bound charges must be uncovered; therefore, the width of the depletion region depends on the density of doping. Under forward bias, the junction potential is reduced and the depletion region narrows; under reverse bias, the depletion region widens.

If the reverse bias is increased sufficiently, a sudden increase in reverse current is observed (see Fig. 11.8).† This behavior is due to the *Zener* effect or the *avalanche* effect. In Zener breakdown, the electric field in the junction becomes high enough to pull electrons directly out of covalent bonds. The electron-hole pairs thus created then contribute a greatly increased reverse current. The avalanche effect occurs at voltages higher than Zener breakdown voltages. At these high voltages, carriers gain sufficient energy between collisions to knock electrons from covalent bonds; these too can reach ionizing energies in a cumulative process.

In either type of breakdown, the reverse current is large and nearly independent of voltage. If the power dissipated is within the capability of the diode, there is no damage; reduction of the reverse bias voltage below V_b reduces the current to I_s. By controlling doping densities, it is possible to manufacture Zener diodes† with breakdown voltages from a few volts to several hundred. These are valuable because of their ability to maintain a nearly constant voltage under conditions of widely varying current.

---

† The alloyed silicon diode was invented by Gerald Pearson in 1952. On his way out to lunch one day, he suggested to his assistant at Bell Telephone Laboratories that he try to alloy electrodes to pure silicon crystals using a variety of metals. When Pearson returned, he discovered that the aluminum specimen exhibited the startling behavior of Fig. 11.8. He first ascribed the reverse current to the Zener effect, but it turned out to be what we now call *avalanche breakdown*.

Tunnel Diodes

If the doping concentrations are increased, thereby increasing the concentration of uncovered charges, the width of the depletion region, and the associated potential barrier, is decreased. If the doping is greatly increased so that the depletion region is less than 100 Å wide, a new conduction mechanism is possible and the device characteristics are radically changed.

As explained by Leo Esaki in his 1958 announcement, for very thin potential barriers quantum mechanics theory indicates that there is a finite probability that an electron may *tunnel* through the barrier without ever possessing enough energy to climb over it. The *I-V* characteristics of an *Esaki diode* are shown in Fig. 11.13. The solid colored line indicates the tunneling effect. At voltages well below the threshold for normal forward current, electrons can tunnel from the *n* region to the *p* region if there are available holes of appropriate energy level into which the electron can move without ever possessing sufficient energy to become free.‡

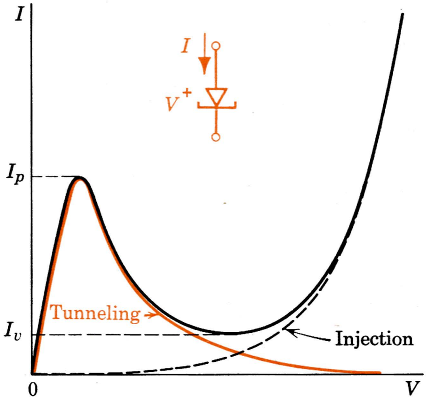


Figure 11.13 Tunnel diode characteristics.

The tunneling current increases with voltage until the effect of forward bias begins to alter the energy relationship and reduce the available holes. After the peak current I_p is reached, the tunneling current *decreases* with increased voltage and the normal injection current begins to dominate.

The peak current I_p and the valley current I_v are stable operating points. Since tunneling is a wave phenomenon, electron transfer occurs at the speed of light and switching between I_p and I_v is fast enough for computer applications. Furthermore, between I_p and I_v there is a region in which the resistance $r = dV/dI$ is *negative*. By offsetting losses in *L* and *C* components, such a negative resistance permits oscillation and the tunnel diode is used as a very high frequency oscillator.

---

† In commercial practice, the term "Zener diodes" includes devices employing both effects.

‡ Millman and Halkias, op. cit.

Metal-Semiconductor Diodes

The leads to a semiconductor diode should make *ohmic contact* to provide for the easy flow of current without creation of additional potential barriers. Aluminum in contact with silicon acts as a p-type impurity and the flow of hole current is easily accomplished by recombination with electrons supplied by the external circuit. Aluminum in contact with n-type material, however, would create a *rectifying contact* instead of the desired ohmic contact. In practice, therefore, contacts to n-type material are made of special alloys or a layer of heavily doped n^+ material is used to provide a low-potential transition between the semiconductor and the metal.

The rectifying contact created between metal and semiconductor differs from a pn junction in two important respects. In such a *Schottky diode* the forward voltage drop is only about half that across a pn junction at the same current. Secondly, since in an Al-n diode there are only majority carriers, switching is very fast because there is no wait for the recombination of injected minority carriers. These characteristics make Schottky diodes useful in integrated circuits.

Photodiodes

In a *photodiode* radiant energy is used to create electron-hole pairs in the region near the junction. Under reverse-biased conditions, the light-injected minority carriers cross the junction and augment the normal reverse saturation current due to thermally generated carriers. The resulting current (a fraction of a milliampere) is approximately proportional to the total illumination and the ratio of "illuminated" current to "dark" current is very high. These characteristics are desirable in a phototransducer. The use of a photodiode in energy conversion is discussed in Chapter 20.

Light-Emitting Diodes

The creation of electron-hole pairs is a reversible process; energy is released when an electron recombines with a hole. In silicon and germanium, recombination usually occurs at defects in the crystal which can *trap* a moving electron or hole, absorb its energy, and hold it until a recombination partner comes along. Only occasionally in silicon or germanium but frequently in a *III-V* compound such as gallium arsenide, an electron drops directly into a hole and a photon of energy is generated. Gallium arsenide junctions providing optimum conditions for the generation of radiation in the visible range are called *light-emitting diodes* (LED). Under special conditions, the light emitted is coherent and the device is a *junction laser.*

SUMMARY

- In a metal, electrons of mobility μ and drift velocity u constitute a current density $J = neu = ne\mu\mathcal{E} = \sigma\mathcal{E}$, defining conductivity σ.
- In a semiconductor, thermally generated electrons and holes constitute a drift current density $J = (n\mu_n + p\mu_p)e\mathcal{E} = \sigma\mathcal{E}$.
  A hole is considered to be a positively charged particle.
- Low-density doping with a pentavalent (trivalent) element produces an n-type (p-type) semiconductor of greatly increased conductivity that is highly temperature dependent.
- For equilibrium, generation rate equals recombination rate and $np = n_i^2$

  In neutral n-type material, $n_n \cong N_d$ and $p_n \cong n_i^2/N_d$.
  In neutral p-type material, $p_p \cong N_a$ and $n_p \cong n_i^2/N_a$.
- The diffusion current density due to nonuniform concentrations of randomly moving electrons and holes is

$$J = J_n + J_p = eD_n \frac{dn}{dx} - eD_p \frac{dp}{dx}$$

- A pn junction diode is a single crystal with an abrupt change from p- to n-type material at the junction.
  Diffusion of majority carriers across the junction uncovers bound charge, creates a potential hill, and encourages an opposing drift of minority carriers.
  Forward biasing reduces the potential hill and encourages diffusion.
  Reverse biasing increases the potential hill and discourages diffusion.

$$I = I_s(e^{eV/kT} - 1)$$

- Normal reverse current is very small and nearly constant.
  At breakdown, reverse current is large and independent of voltage.
- Heavily doped Esaki tunnel diodes provide valuable characteristics.
  Schottky metal-semiconductor diodes have special applications.
  Photodiodes use radiant energy to create electron-hole pairs.
  Light-emitting diodes use the energy released in direct recombination.

REVIEW QUESTIONS

1. Explain why drift velocity in a solid is proportional to voltage V, whereas in a vacuum u is proportional to $V^{1/2}$.
2. How does resistance vary with temperature for a conductor? For a semiconductor? Why?
3. How does the speed of transmission of a "dot" along a telegraph wire compare with the drift velocity?
4. Define the following terms: mobility, conduction band, covalent bond, electron-hole

generation, recombination, doping, and intrinsic.

5. Describe the formation of a hole and its role in conduction.
6. In a semiconductor, what is the effect of doping with donor atoms on electron density? On hole density?
7. Is a semiconductor negatively charged when doped with donor atoms?
8. Why does the factor eV/kT appear in thermionic emission, electron-hole generation, and *pn* junction current flow?
9. How do you expect the term n_i^2 to vary with temperature?
10. Is the approximation of Eq. 11-12 justified if $N_d = 100n_i$?
11. What are the two charge carriers and the two conduction mechanisms present in semiconductors?
12. Cite two nonelectrical examples of diffusion.
13. Define the following terms: junction, uncovered charge, majority carriers, transition region, injection, and potential hill.
14. Describe the forward, reverse, and breakdown characteristics of a junction diode.
15. Describe with sketches the operation of an *np* diode.
16. How is contact potential utilized in a thermocouple?
17. Sketch curves near the origin to distinguish between the v-i characteristics of ideal, thermionic, and semiconductor diodes.
18. Why does reverse saturation current vary with temperature?
19. Describe the behavior of a Zener diode under forward and reverse bias.

EXERCISES

1. Stating any necessary assumptions, show that the spacing between atoms in copper and germanium crystals is on the order of 1 Å.
2. If electrons drift at a speed of 0.5 mm/s in ordinary residential wiring (No. 12 copper wire, diameter = 0.081 in.), what current does this represent?
3. Given that resistance $R = \rho l/A$, derive an expression for resistivity ρ in terms of carrier density, mobility, and charge.
4. The average energy of a crystal at temperature T is indicated by kT.
   (a) Show that eV and kT have the same units and calculate e/kT for room temperature 20°C.
   (b) Sketch a graph of $e^{-eV/kT}$ for 20°C and $0.2 < V < 2$ and write a brief statement describing the relation between thermal generation rate and energy gap.
5. For $V = 1$ V, sketch a graph of $e^{-eV/kT}$ for $200 < T < 400$ K and write a brief statement describing the relation between thermal generation rate and temperature.
6. A silicon sample contains boron at a concentration of 2×10^{20} atoms/m^3.
   (a) Estimate the hole and electron concentrations at room temperature. Is this *p*-type or *n*-type material?
   (b) Repeat part (a) for 300°C where $n_i = 3 \times 10^{21}/\text{m}^3$.
7. Estimate the relative concentration of germanium atoms and electron-hole pairs at room temperature, and predict the intrinsic resistivity.
8. A silicon crystal is doped with 10^{20} aluminum atoms/m^3.
   (a) Justify *quantitatively* the statement that: "The metallurgical properties of the doped crystal are not significantly altered but the electrical properties are greatly changed."
   (b) Determine the factor by which the conductivity has been changed by doping.
9. A pure silicon chip has a resistance of 40 Ω. The resistance is to be reduced to 0.4 Ω by doping with trivalent boron. Specify the doping density and identify the important charge carriers.
10. A chip of intrinsic silicon is 1 mm × 2 mm in area and 0.1 mm thick.
    (a) What voltage is required for a current of 2 mA between opposite faces?
    (b) Repeat part (a) for a current of 2 mA between opposite 0.1 × 1 mm ends.
11. The block of Exercise 10 is doped by adding 1 atom of phosphorus per 10^7 atoms of silicon.

(a) How many atoms per hole-electron pair are there in intrinsic silicon?
(b) Is the doped material *n*-type or *p*-type? Why?
(c) In the doped material, what is the density of majority and minority carriers? Can one be neglected?
(d) What voltage is required for a current of 2 mA between faces?
(e) By what factor has doping increased the conductivity?

12. Repeat Exercise 11, assuming that 1 in every 10^8 atoms of silicon is replaced by an atom of boron.

13. A certain block of pure silicon has a resistance of 200 Ω. It is desired to reduce this resistance to 0.2 Ω by doping with pentavalent arsenic.
(a) What are the charge carriers of importance in the doped semiconductor?
(b) What fraction of the silicon atoms must be replaced by arsenic atoms?

14. The resistivity of an *n*-type silicon specimen is specified to be 0.2 Ω-m. Estimate the electron and hole concentrations.

15. Consider three specimens: a pure metal, a pure semiconductor, and a moderately doped semiconductor, at room temperature. Predict the effect on the conductivity of each specimen by a 10% increase in absolute temperature (K). Write an appropriate expression for conductivity for each; predict the new conductivity by selecting the most appropriate phrase from the following:
(a) about the same
(b) increased by about 10%
(c) increased by more than 10%
(d) decreased by about 10%
(e) decreased by more than 10%
and explain, briefly, your reasoning.

16. A 5000-Ω resistor is to be fabricated (in an "integrated circuit") as a narrow strip of *p*-type silicon 4 μm thick. If the strip is 25 μm wide and 600 μm long, what concentration of acceptor atoms is required?

17. (a) Calculate the diffusion constants D_n and D_p for silicon at room temperature.
(b) At a junction created by heating an *n*-type silicon crystal in the presence of boron gas (see Fig. 11.7c), the density of electrons drops from 6×10^{20} to 5×10^{20} per m^3 in a distance 0.02 μm. Estimate the diffusion current density across the junction.

18. In a germanium *pn* junction diode, the density of holes drops from 10^{21} to 0.9×10^{21}/m^3 in 2 μm. Estimate the diffusion current density due to holes across the junction at room temperature.

19. Draw a "rectifier" circuit and describe, with sketches, how an *np* junction diode can be used to rectify alternating currents.

20. A diode consists of a left-hand portion doped with 10^{20} antimony atoms/m^3 and a right-hand portion doped with 4×10^{20} boron atoms/m^3.
(a) Sketch the diode, indicating the junction, and identifying the minority and majority carriers in each portion.
(b) For electrical neutrality, how do the widths of the depletion layers in *n* and *p* regions compare?
(c) Sketch a graph of charge density across the unbiased diode.

21. A semiconductor diode consists of a left-hand portion doped with arsenic and a right-hand portion doped with indium.
(a) Sketch the diode, indicating the junction. Below the sketch, draw graphs of charge density, electric field, and potential distribution across the unbiased diode.
(b) Draw labeled arrows indicating directions of drift and diffusion of electrons, holes, and net currents.
(c) For forward bias, indicate the necessary polarity.

22. In a reverse-biased junction, when all the available charge carriers are being drawn across the junction, increasing the reverse junction voltage does not increase the reverse current.
(a) At what voltage will the reverse current in a germanium *pn* junction at room temperature reach 98% of its saturation value?
(b) If the reverse saturation current is 20 μA, calculate the current for a forward bias voltage of the magnitude determined in part (a).

23. For a silicon diode at room temperature, the manufacturer specifies a reverse saturation current of 1 nA and a forward current of 40 mA at a forward voltage of 1 V.
(a) Estimate the resistance of the body of

the diode, the so-called *bulk resistance.*

(b) Predict the forward voltage required for a current of 100 mA.

(c) Repeat part (b) for a current of 2 mA.

24. For a germanium diode carrying 10 mA, the required forward bias is about 0.2 V. Estimate the reverse saturation current and the bias voltages required for currents of 1 and 100 mA, respectively.

25. The diode of Exercise 23 is connected in series with a 150-Ω resistor and a 3-V battery. Stating any necessary assumption, predict the current.

26. A Zener diode with a breakdown voltage of 3 V has a reverse saturation current of 5 μA; the internal (body) resistance is 2 Ω. Sketch the v-i characteristics for $-20 < i < +20$ mA.

PROBLEMS

1. Sketch in perspective a rectangular block of semiconductor of unknown type in a magnetic field B out of the paper. When a current I is introduced from right to left, it is observed that the upper surface exhibits a positive potential with respect to the lower (the Hall effect).

(a) Is the specimen p-type or n-type semiconductor?

(b) How does this experiment justify the concept of the hole as a mobile positive charge?

2. If donor impurity atoms are added to an intrinsic semiconductor, how is the hole density affected? If 10^{16} atoms/m^3 of pentavalent antimony are added to pure silicon, predict the new hole density.

3. An experimental melt contains 100 g of pure germaniun, 5×10^{-5} g of indium, and 3×10^{-5} g of antimony. Determine the conductivity of a sample of this germanium alloy.

4. For a nonlinear device, the incremental resistance is defined as $r = dv/di$. For an ideal semiconductor diode, derive a general expression for r in terms of I_s and an approximate expression in terms of I the forward current. Evaluate r for a reverse current of 5 μA and for a forward current at 5 mA.

5. Using your understanding of Zener and avalanche breakdown mechanisms and your knowledge of the effect of temperature on the energies of electrons and lattice atoms, predict (qualitatively) the effect of an increase in temperature on the breakdown voltage V_b for diodes governed by each of the two mechanisms.

Chapter 12
Diode Applications

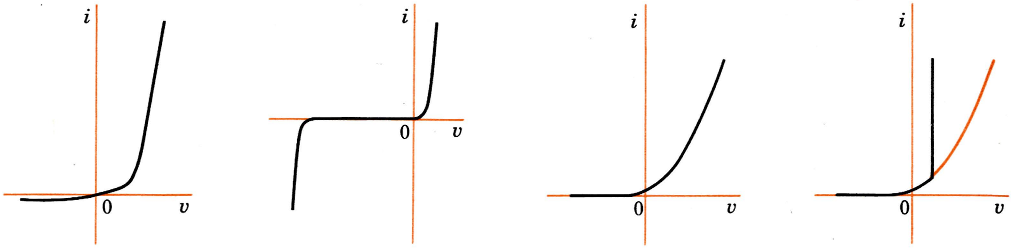
Diode Models
Rectifiers and Filters
Waveshaping Circuits

Using ideal models of practical circuit elements, we developed tools for the analysis of linear networks under a variety of circumstances. Electronic devices, however, are inherently nonlinear. For example, in a space-charge controlled vacuum tube, the anode current varies as the three-halves power of the electrode voltages, and in a semiconductor the current across a junction varies exponentially with the junction voltage. It is possible to analyze the performance of such devices graphically as in Chapter 10, but this is laborious and the results lack generality. Fortunately, under certain conditions electronic devices can be represented by linear circuit models to which the methods of linear analysis can be applied. In working with diodes, relatively simple models are satisfactory.

After we see how to represent diodes by appropriate models, we learn to predict the performance of diode devices and, in simple cases, to design circuits to meet specifications. Our approach is to study a few of the more common applications and to outline principles that are generally useful. The diode applications selected for discussion include rectifiers and filters, voltage regulators, and waveshaping circuits.

Diode Models

The diode is a two-terminal, nonlinear device that presents a relatively low resistance to current flow in one direction and a relatively high resistance in the other. In a semiconductor diode (Fig. 12.1a), under conditions of forward bias the diffusion of majority carriers is facilitated and an appreciable current will flow with a junction voltage of a few tenths of a volt; with reverse bias, there is only the current due to the motion of thermally generated minority carriers and this is very small. In a Zener diode (Fig. 12.1b), when the reverse bias reaches the breakdown value the reverse current increases rapidly with only a small increase in voltage.



(a) Semiconductor (b) Zener (c) Vacuum (d) Gas

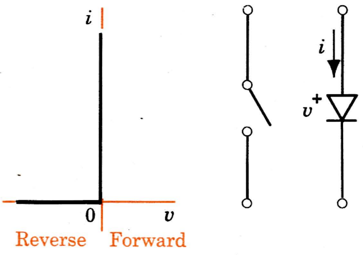
Figure 12.1 Diode characteristics.

In a vacuum diode with zero applied voltage, there is a small current due to the few electrons emerging from the cathode with initial velocities high enough to overcome the space charge (Fig. 12.1c). With a small negative voltage, the current is zero; for positive voltages the current increases rapidly in the space-charge limited region. In a gas diode, when the voltage reaches the ionizing potential of the gas there is a rapid increase in current with no further increase in voltage, and the characteristics are as in Fig. 12.1d.

Circuit Models

The electrical model of a device is frequently called a *circuit model* because it is composed of ideal circuit elements. The models are derived from the characteristic curves of the real devices and, when properly devised, can be used to predict accurately the behavior of complicated vacuum-tube and semiconductor devices in practical applications.

The important diode characteristic is the discrimination between forward and reverse voltages. The ideal diode presents no resistance to current flow in the forward direction and an infinite resistance to current flow in the reverse direction. A selective switch, which is closed for forward voltages and open for reverse voltages, is equivalent to the ideal diode since it has the same current-voltage characteristic. Such a switch is represented in circuit diagrams by the symbol shown in Fig. 12.2c; the triangle points in the direction of forward current flow.



(a) (b) (c)

Figure 12.2 Ideal diode representation.

The vacuum diode approaches the ideal in the reverse direction, but it presents a finite resistance in the forward direction. Its behavior can be approximated by a circuit model based on the linearized characteristic curve of Figure 12.3. For reverse voltages, the switch is open; for forward voltages, the switch is closed and the device presents a resistance R_F to current flow. In the case shown, the forward resistance is an average value estimated from the slope of the linearized curve. Here (and in Example 1 on p. 313),

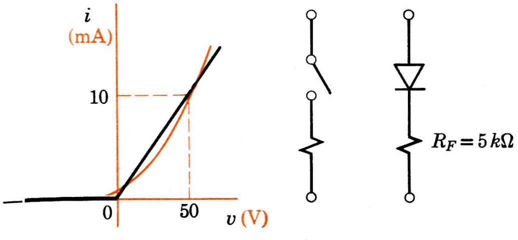


Figure 12.3 Circuit models for a vacuum diode.

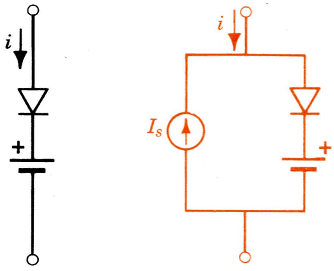


Figure 12.4 A semiconductor diode.

$$R_F = \frac{V}{I} = \frac{50\ \text{V}}{0.01\ \text{A}} = 5000\ \Omega$$

In a semiconductor diode, only a few tenths of a volt (0.3 V for germanium, 0.7 V for silicon) of forward bias is required for appreciable current flow. The combination of an ideal diode and a voltage source shown in Fig. 12.4a is usually satisfactory for predicting diode performance. If the reverse saturation current in a semiconductor diode is appreciable, this fact must be incorporated into the model. One possibility is to use a current source with a magnitude I_s as shown in Fig. 12.4b.†

Example 1

The diode of Fig. 12.3 is connected in the circuit of Fig. 12.5a. Predict the current in the load resistor $R_L = 5\ \text{k}\Omega$ if $v = 100 \sin \omega t$ V.

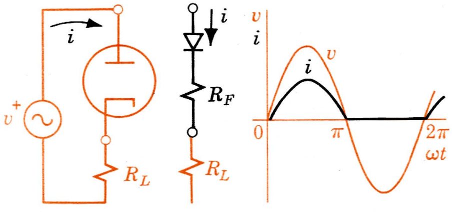


Figure 12.5 Use of diode models.

The vacuum diode is replaced by the circuit model shown in Fig. 12.5b. For $0 < \omega t < \pi$, v is positive, corresponding to forward bias, and

$$i_L = \frac{v}{R_F + R_L} = \frac{100 \sin \omega t}{(5 + 5) \times 10^3} = 10 \sin \omega t \text{ mA}$$

For $\pi < \omega t < 2\pi$, v is negative, corresponding to negative bias, and no current flows.

The answer is

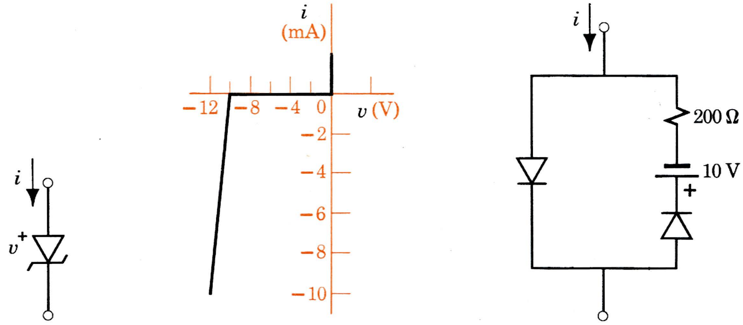
$$\begin{cases} i_L = 10 \sin \omega t \text{ mA} & \text{for } 0 < \omega t < \pi \\ i_L = 0 & \text{for } \pi < \omega t < 2\pi \end{cases}$$

as sketched in Fig. 12.5c.

---

† For example, the manufacturers' ratings for the IN4004 subminiature silicon rectifier indicate conservative values as follows: 1 V drop at 1 A in the forward direction and only 10 μA reverse current under rated conditions.

The characteristic curve for a Zener diode can be linearized as in Fig. 12.6. When forward biased (v positive), current flows freely; the forward resistance is small and can be neglected. For reverse bias voltages greater than the breakdown potential, the resistance is estimated to be



(a) Symbol (b) Characteristics (c) Model

Figure 12.6 Zener diode representation.

$$R_Z = \frac{\Delta v}{\Delta i} = \frac{12 - 10}{0.01 - 0} = 200\ \Omega$$

The circuit model includes a voltage source to indicate that reverse current does not flow until the negative voltage across the diode exceeds 10 V.

Example 2

The Zener diode of Fig. 12.6 is connected in the circuit of Fig. 12.7a.

Represent the diode in the breakdown region by an appropriate model and redraw the circuit.

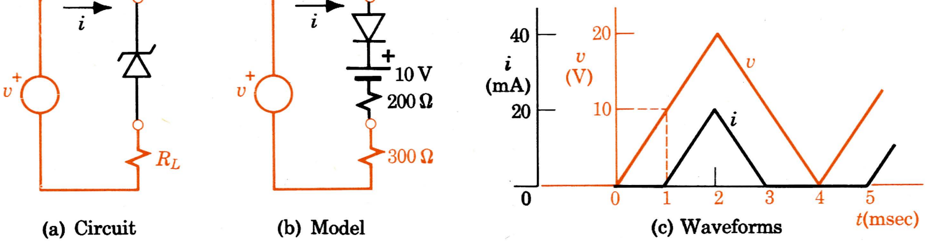
For the triangular voltage waveform shown, predict the current i.

For this connection, the circuit model is as shown in Fig. 12.7b. For $v > 10$ V,

$$i = \frac{v - 10}{R_Z + R_L} = \frac{v - 10}{200 + 300}$$

The maximum value, at $t = 2$ ms (Fig. 12.7c), is

$$I_{\max} = \frac{20 - 10}{500} = 20\text{ mA}$$



(a) Circuit (b) Model (c) Waveforms

Figure 12.7 Zener diode application.

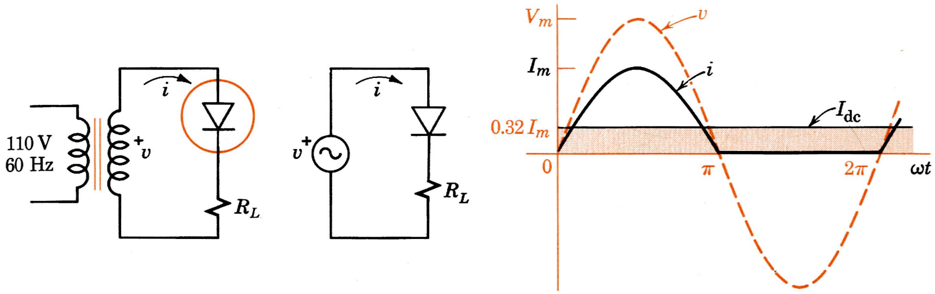
Rectifiers

The nonlinear characteristic of a diode is used to convert alternating current into unidirectional, but pulsating, current in the process called *rectification.* The pulsations are removed in a frequency-selective circuit called a *filter.* Rectifier circuits employ one, two, or four diodes to provide various degrees of rectifying effectiveness. Filter circuits use the energy storage capabilities of inductors and capacitors to smooth out the pulsations and to provide a steady output current. A combination of rectifier, filter, and regulator is called a *power supply.*

Half-Wave Rectifier

Example 1 (Fig. 12.5) illustrates *half-wave rectification.* Ideally, a diode should conduct current freely in the forward direction and prevent current flow in the reverse direction. Practical diodes only approach the ideal. Semiconductor diodes, for example, present a small but appreciable voltage drop in the forward direction and permit a finite current to flow in the reverse direction. For most calculations, the reverse current flow is negligibly small and the forward voltage drop can be neglected with little error.

A practical rectifier circuit is shown in Fig. 12.8a. A transformer supplied from 110-V, 60-Hz house current provides the desired output voltage, which is applied to a series combination of diode (a semiconductor diode is shown) and load resistance R_L. For approximate analysis, the actual diode is represented by an ideal diode and V_F is neglected; the internal resistance of the transformer is neglected also. For $v = V_m \sin \omega t$, the resulting current is



(a) Wiring diagram (b) Circuit model (c) Output waveform

Figure 12.8 A half-wave rectifier.

$$\begin{cases} i = \dfrac{v}{R_L} = \dfrac{V_m \sin \omega t}{R_L} & \text{for } 0 \le \omega t \le \pi \\ i = 0 & \text{for } \pi \le \omega t \le 2\pi \end{cases} \tag{12-1}$$

as shown in Fig. 12.8c.

The purpose of rectification is to obtain a unidirectional current. The dc component of the load current is the average value or

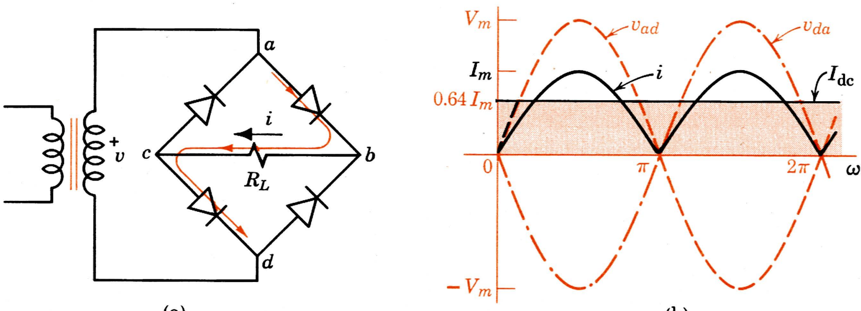
$$I_{\mathrm{dc}} = \frac{1}{2\pi}\int_0^{2\pi} i\,d(\omega t) = \frac{1}{2\pi}\int_0^{\pi} \frac{V_m \sin \omega t}{R_L}\,d(\omega t) + 0$$

$$= \frac{1}{2\pi}\frac{V_m}{R_L}\Big[-\cos \omega t\Big]_0^{\pi} = \frac{V_m}{\pi R_L} = \frac{I_m}{\pi} \tag{12-2}$$

The current through the load resistance consists of half-sinewaves, and the dc component is approximately 30% of the maximum value.

Full-Wave Rectifier

The bridge rectifier circuit of Fig. 12.9 provides a greater dc value from the same transformer voltage. When the transformer voltage $v = v_{ad}$ is positive, the current flow is along path *abcd* as shown and a half-sinewave of current results. When the applied voltage reverses, the voltage $v_{da} = -v_{ad}$ is positive and the current flow is along path *dbca*. The current through the load resistance is always in the same direction, and the dc component is twice as large as in the half-wave rectifier or

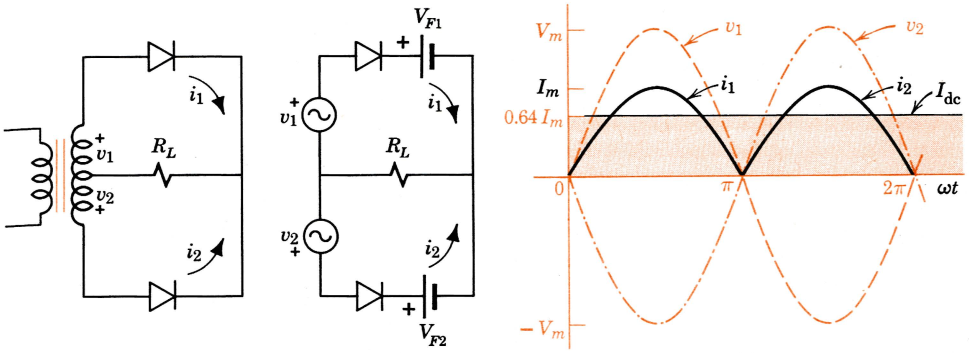


(a) (b)

Figure 12.9 A full-wave bridge rectifier.

$$I_{\mathrm{dc}} = \frac{2}{\pi}\frac{V_m}{R_L} = \frac{2I_m}{\pi} \tag{12-3}$$

The bridge circuit is disadvantageous because four diodes are required, and two diodes and their power-dissipating voltage drops are always in series with the load. The full-wave rectifier circuit of Fig. 12.10 uses a more expensive transformer to produce the same result with only two diodes and with higher operating efficiency. The second output winding on the transformer provides a voltage v_2 that is 180° out of phase with v_1; such a *center-tapped* winding serves as a *phase inverter.* While v_1 is positive, current i_1 is supplied through diode 1; while v_1 is negative, no current flows through diode 1 but v_2 is positive



(a) (b) (c)

Figure 12.10 A full-wave rectifier with phase inverter.

and, therefore, current i_2 is supplied through diode 2. The current through the load resistance is $i_1 + i_2$ and $I_{dc} = 2I_m/\pi$ as in the bridge circuit.

In addition to the forward resistance R_F, forward voltage drop V_F, and reverse current I_s, the important design parameters of a diode are the maximum forward current and the *peak inverse voltage.* For example, diffused-junction silicon rectifiers are available in ratings up to 600 A of rectified forward current. The peak inverse voltage (PIV) is the maximum reverse bias that can be applied without breakdown and it may be critical in high-voltage power supplies. Silicon rectifiers are available in PIV ratings up to 1000 V; for higher voltage applications, rectifiers can be connected in series if provision is made to ensure equal voltage division between units.

Ripple Factor

The desired result of rectification is direct current, but the output currents of the rectifier circuits described obviously contain large alternating components along with the dc component. As a measure of the effectiveness of rectification we define the *ripple factor r* where

$$r = \frac{I_{ac}}{I_{dc}} = \frac{V_{ac}}{V_{dc}} = \frac{\text{rms value of ac components}}{\text{dc component}} \tag{12-4}$$

Since the power dissipated in the load resistance defines the rms value of current, and since the total power is the sum of the power dissipated by the direct and alternating components (see Problem 3),

$$I_{rms}^2 R_L = I_{dc}^2 R_L + I_{ac}^2 R_L \tag{12-5}$$

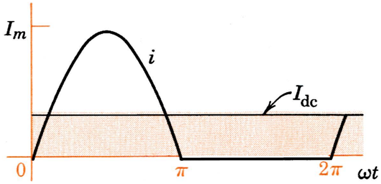
Therefore, $I_{ac}^2 = I_{rms}^2 - I_{dc}^2$ and

$$r = \frac{\sqrt{I_{rms}^2 - I_{dc}^2}}{I_{dc}} = \sqrt{\left(\frac{I_{rms}}{I_{dc}}\right)^2 - 1} \tag{12-6}$$

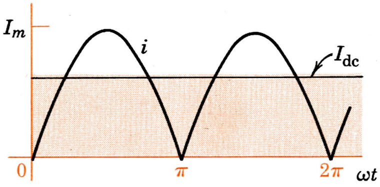
If the ripple factor is low, the circuit is performing the conversion from alternating current to direct current effectively.†

Example 3

Determine and compare the ripple factors for the outputs of the half-wave and full-wave rectifier circuits (Fig. 12.11).



(a) Half-wave output



(b) Full-wave output

Figure 12.11 Ripple factor calculations.

For the half-wave output,

$$I_{\text{rms}} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} i^2\, d(\omega t)}$$

$$= \sqrt{\frac{1}{2\pi}\int_0^{\pi} (I_m \sin \omega t)^2\, d(\omega t)}$$

$$= I_m\sqrt{\frac{\pi/2}{2\pi}} = \frac{I_m}{2} \tag{12-7}$$

Therefore (by Eqs. 12-2 and 12-7),

$$\frac{I_{\text{rms}}}{I_{\text{dc}}} = \frac{I_m/2}{I_m/\pi} = \frac{\pi}{2} = 1.57$$

and

$$r = \sqrt{(1.57)^2 - 1} = 1.21 \text{ or } 121\%$$

For the full-wave output,

$$I_{\text{rms}} = I_m\sqrt{\pi/2\pi} = I_m/\sqrt{2} \quad \text{and} \quad I_{\text{dc}} = 2I_m/\pi$$

$$\frac{I_{\text{rms}}}{I_{\text{dc}}} = \frac{I_m/\sqrt{2}}{2I_m/\pi} = \frac{\pi}{2\sqrt{2}} = 1.11$$

and

$$r = \sqrt{(1.11)^2 - 1} = 0.48 \text{ or } 48\%$$

Filters

Although using full-wave instead of half-wave rectification reduces the ac component from 121% to 48% of the dc component, the output is still unsatisfactory for most electronic purposes.

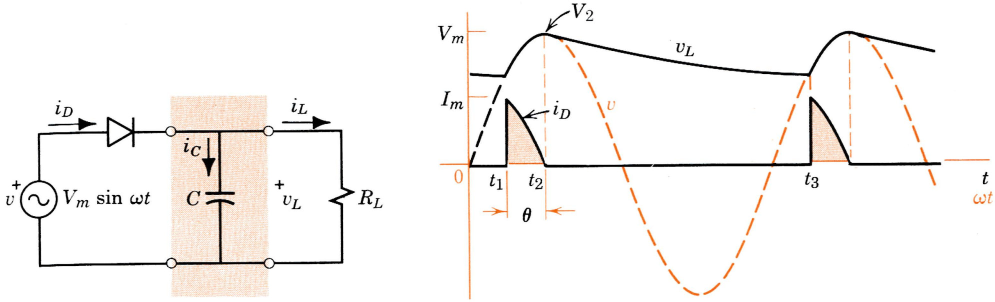
Capacitor Filter

The ripple factor can be greatly reduced by a *filter* consisting of a capacitor shunted across the load resistor. The capacitor can be thought of as a low-impedance path taken by the ac components of the rectified wave, or it can be thought of as a "tank" that stores charge during the period when the diode is conducting and releases charge to the load during the nonconducting period.

---

† For a more complete discussion of rectifiers and filters, see Millman and Halkias, *Electronic Devices and Circuits*, McGraw-Hill Book Co., New York, 1967.

If the diode resistance is small and if the steady state has been reached, the operation is as shown in Fig. 12.12. At time $t = 0$, the source voltage v is zero



(a) Circuit

(b) Voltage and current waveforms

Figure 12.12 A capacitor filter.

but the load voltage $v_L = v_C$ is appreciable because the previously charged capacitor is discharging through the load. At $t = t_1$, the increasing supply voltage v slightly exceeds v_L and the diode conducts. The diode current i_D rises abruptly to satisfy the relation $i_C = C\,dv/dt$ and then decreases to zero; the diode switches off when v drops below v_L. During the charging period, $t_1 < t < t_2$,

$$v_L = V_m \sin \omega t \tag{12-8}$$

During the discharging period, $t_2 < t < t_3$,

$$v_L = V_2\, e^{-(t-t_2)/RC} \tag{12-9}$$

as expected from the natural response of the RC circuit. At time t_3, the supply voltage again exceeds the load voltage and the cycle repeats.

The load current i_L is directly proportional to load voltage v_L. Because i_L never goes to zero, the average value or dc component is relatively large as compared to the half-wave rectifier alone and the ac component is correspondingly lower. The ripple factor is greatly reduced by the use of the capacitor.

Capacitor Filter—Approximate Analysis

Design charts are available in handbooks published by rectifier manufacturers[†] that relate R_L, C, r, and V_L/V_m. The following approximate analysis, which gives satisfactory results for most purposes, illustrates the roles played by the various circuit parameters. We assume that:

1. The time constant R_LC is large enough so that the charging interval $t_2 - t_1$ is small compared to time T for one cycle.

[†] Page 135, *Zener Diode and Rectifier Handbook*, Motorola Inc., Phoenix, Arizona, 1961.

2. The diode current i_D, a portion of a cosine wave, can be approximated by a triangular pulse.
3. Diode switching time t_2 occurs at the peak of the supply voltage wave so that $V_2 = V_m$.
4. The sinusoidal and exponential segments of the v_L curve can be approximated by straight lines.

On the basis of these assumptions, the behavior is described by Fig. 12.13.

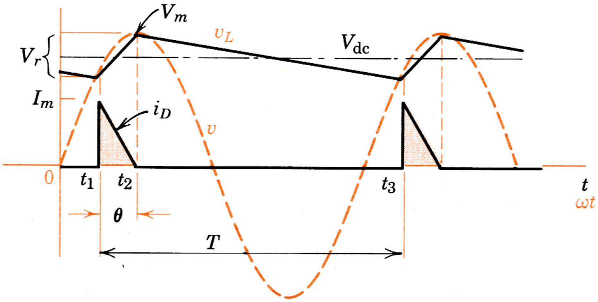


Figure 12.13 Approximate analysis of a capacitor filter.

For a given load and capacitor, we are interested in determining the necessary supply voltage, the diode rating, and the resulting voltage variation. If the charging interval is negligibly small, the load current is supplied by the capacitor and the charge transferred is

$$Q = I_{dc}T \cong C\,\Delta V_C = CV_r \tag{12-10}$$

Solving, the ripple voltage is

$$V_r \cong \frac{I_{dc}T}{C} = \frac{I_{dc}}{fC} \tag{12-11}$$

where f is the frequency, usually 60 Hz. The necessary supply voltage is

$$V_m = V_{dc} + \frac{V_r}{2} = V_{dc} + \frac{I_{dc}}{2fC} \tag{12-12}$$

The charging interval for the capacitor corresponds to the *conduction angle* for the diode. Since $V_m - V_r = V_m \sin(90° - \theta) = V_m \cos\theta$,

$$\theta = \cos^{-1}\frac{V_m - V_r}{V_m} \tag{12-13}$$

and the conduction *time* is

$$t_2 - t_1 = \frac{\theta}{2\pi}T \tag{12-14}$$

The charge carried through the load is equal to the charge conducted by the diode during the triangular pulse; therefore

$$I_{\mathrm{dc}}T \cong \tfrac{1}{2}I_m(t_2 - t_1) = \tfrac{1}{2}I_m\frac{\theta}{2\pi}T$$

or

$$I_m \cong I_{\mathrm{dc}}\frac{4\pi}{\theta} = I_{\mathrm{dc}}\frac{720^\circ}{\theta^\circ} \tag{12-15}$$

The ripple voltage consists of a triangular wave of maximum value $V_r/2$ and rms value $V_r/2\sqrt{3}$ (see Exercise 3-19b). The ripple factor is

$$r = \frac{V_{\mathrm{ac}}}{V_{\mathrm{dc}}} = \frac{V_r}{2\sqrt{3}V_{\mathrm{dc}}} = \frac{I_{\mathrm{dc}}/V_{\mathrm{dc}}}{2\sqrt{3}fC} = \frac{1}{2\sqrt{3}fCR_L} \tag{12-16}$$

These relations apply to a half-wave rectifier; similar equations can be derived for a full-wave rectifier. By using these approximate relations, the performance of an existing filter can be determined or a filter can be designed to meet specifications as in Example 4.

Example 4

A load ($R_L = 4000\ \Omega$) is to be supplied with 200 V at 50 mA with a ripple factor of less than 2%. Design a rectifier-filter combination to meet these specifications.

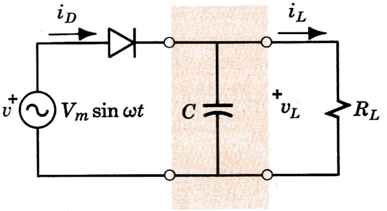


Figure 12.14 Approximate design of a rectifier-filter circuit.

Assuming a 60-Hz supply and the half-wave rectifier with capacitor filter of Fig. 12.14, the design is

$$C = \frac{1}{2\sqrt{3}rfR_L} = \frac{1}{2\sqrt{3} \times 0.02 \times 60 \times 4000} = 60\ \mu\mathrm{F}$$

$$V_r = 2\sqrt{3}V_{\mathrm{dc}}r = 2\sqrt{3} \times 200 \times 0.02 = 13.9\ \mathrm{V}$$

$$V_m = V_{\mathrm{dc}} + \frac{V_r}{2} = 200 + \frac{13.9}{2} \cong 207\ \mathrm{V}$$

$$\theta = \cos^{-1}\frac{V_m - V_r}{V_m} = \cos^{-1}\frac{193}{207} = 21.5^\circ$$

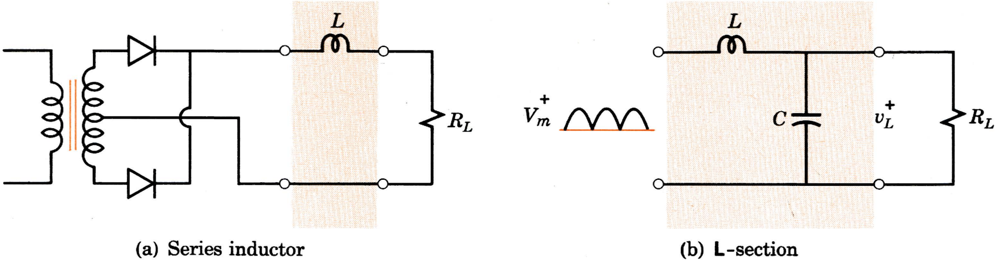
$$I_m = I_{\mathrm{dc}}\frac{720^\circ}{\theta^\circ} = 0.050 \times \frac{720}{21.5} = 1.68\ \mathrm{A}$$

The peak inverse voltage occurs when the supply voltage reaches its negative maximum and the voltage across the diode is $V_m + v_L \cong 2V_m = 414$ V.

A diode rated at $I_m = 1.75$ A and PIV $= 500$ V would be satisfactory.

L-*Section Filters*

Instead of a capacitor in parallel with the load resistance, an inductor or "choke" coil in series with the load can be used to reduce the ripple factor (Fig. 12.15a). An inductance offers relatively high impedance to the ac components



(a) Series inductor (b) **L**-section

Figure 12.15 Filters employing inductors.

of current while passing the dc components without opposition. However, inductors in the sizes required for low ripple are relatively expensive; for a given investment, the combination of inductance and capacitance shown in Fig. 12.15b gives better results. Such a two-port is called an **L**-*section* because of its geometry.

If the inductance L is greater than a certain critical value† L_c where

$$L_c \cong \frac{R_L}{6\pi f} \cong \frac{R_L}{1000} \tag{12-17}$$

for 60-Hz operation, the flywheel effect of the inductance keeps the current flowing continuously (instead of in short pulses) and the output voltage changes only slightly with changes in R_L. The analysis of the filtering action under this condition is a straightforward application of ac circuit theory.

As represented in Fig. 12.15b, the output of a full-wave rectifier consists of a series of half-sinewaves. The Fourier series representation of such a periodic wave is

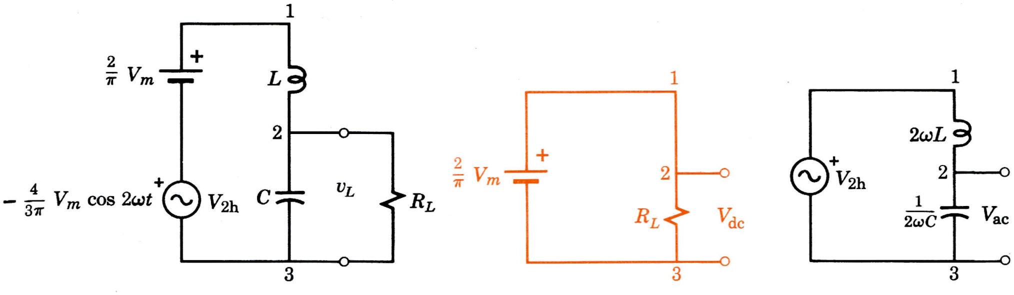
$$v = \frac{2}{\pi} V_m(1 - \tfrac{2}{3} \cos 2\omega t - \tfrac{2}{15} \cos 4\omega t - \tfrac{2}{35} \cos 6\omega t - \cdots) \tag{12-18}$$

The first term, a constant, represents the dc component of relative magnitude 1. The second term is a second harmonic of relative magnitude $\frac{2}{3}$ and frequency 2ω. The succeeding terms in this infinite series are smaller and, furthermore, they are easier to filter out. Our analysis is concerned with the second harmonic term because if that is filtered out effectively, all the higher harmonics are removed even more effectively (see Exercise 20).

A common problem is to determine the ripple factor in the voltage across R_L for given values of L and C. These values are not critical and it is convenient

---

† See Millman and Halkias, *Electronic Devices and Circuits*, McGraw-Hill Book Co., New York, 1967.



(a) (c)

Figure 12.16 Analysis of an L-Section filter by superposition.

to make simplifying assumptions that reduce the work involved. In the circuit of Fig. 12.16 let us assume that:

1. The input voltage to the filter is $v = (2/\pi)V_m - (4/3\pi)V_m \cos 2\omega t$.
2. The resistance of inductor L is negligibly small and, therefore, the full dc component of voltage appears across R_L.
3. The reactance $1/2\omega C$ is small compared to R_L (as it must be for good filtering) and $Z_{RC} = Z_{23} \cong -j\,1/2\omega C$.
4. The reactance of the inductor $2\omega L$ is large compared to Z_{RC} (as it must be for good filtering), and $Z_{13} \cong j(2\omega L - 1/2\omega C) \cong +j\,2\omega L$.

Since the filter is linear, the principle of superposition is applicable and the circuit of Fig. 12.16a can be replaced by an equivalent combination of two circuits. By inspection of the voltage divider in Fig. 12.16c (R_L has been neglected) and considering magnitudes only,

$$\frac{V_{ac}}{V_{2h}} = \frac{Z_{23}}{Z_{13}} = \frac{1/2\omega C}{2\omega L} \cong \frac{1}{4\omega^2 LC} \tag{12-19}$$

or the output is

$$V_{ac} = \frac{1}{4\omega^2 LC} V_{2h} = \frac{1}{4\omega^2 LC} \frac{4V_m}{3\pi\sqrt{2}}$$

This is the only significant ripple component, so the ripple factor is

$$r = \frac{V_{ac}}{V_{dc}} = \frac{4V_m}{4\omega^2 LC(3\pi\sqrt{2})} \cdot \frac{\pi}{2V_m} = \frac{\sqrt{2}/3}{4\omega^2 LC} = \frac{0.47}{4\omega^2 LC} \tag{12-20}$$

Example 5 on the next page illustrates ripple calculation.

Example 5

The load of Example 4 ($R_L = 4000\ \Omega$) is to be supplied from a full-wave rectifier with an L-section filter consisting of $L = 10$ H and $C = 10\ \mu$F. Check the assumptions made in deriving Eq. 12-20 and determine the ripple factor.

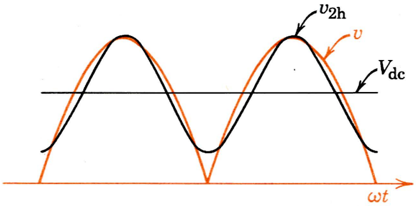


Figure 12.17 Calculating ripple factor from the Fourier components of the filter input.

By Eq. 12-17, $L_c = 4000/1000 = 4$ H. Therefore $L > L_c$ and the ac analysis is applicable.

Assuming 60 Hz operation, for the second harmonic,

$$X_L = 2\omega L = 2(2\pi\, 60)10 = 7540\ \Omega$$

$$X_C = -\frac{1}{2\omega C} = -\frac{1}{2(2\pi\, 60)10^{-5}} = -133\ \Omega$$

and assumptions 3 and 4 appear to be justified.

The rms value of the second harmonic input component in Fig. 12.17 is, by Eq. 12-18,

$$V_{2h} = \frac{2}{3\sqrt{2}} V_{dc} \cong 94\ \text{V}$$

The filter reduces this to (Eq. 12-19):

$$V_{ac} \cong \frac{1/2\omega C}{2\omega L} V_{2h} = \frac{133}{7540} 94 = 1.66\ \text{V}$$

The ripple factor is

$$r = \frac{V_{ac}}{V_{dc}} = \frac{1.66}{200} \cong 0.008 \cong 0.8\%$$

The same result is obtained directly from Eq. 12-20.

To reduce the ripple factor further, a second L-section can be added before the load resistor. For two identical sections, the denominator of Eq. 12-20 becomes $(4\omega^2 LC)^2$.

Zener-Diode Voltage Regulator

Filters are designed to minimize the rapid variations in load voltage due to cyclical variations in rectifier output voltage. There are other reasons for load voltage variation. If the amplitude of the supply voltage (V_m) fluctuates, as it does in practice, the dc voltage will fluctuate. If the load current changes due to a change in R_L, the dc voltage will change due to the IR drop in the transformer, rectifier, and inductor (if any). A filter cannot prevent these types of variation and if the load voltage is critical, a *voltage regulator* must be employed. Gas tubes and Zener diodes are suitable for voltage regulation.

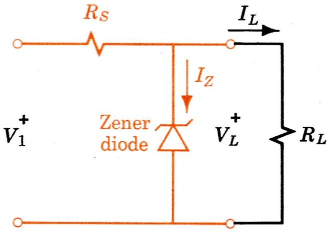


Figure 12.18 A Zener diode voltage regulator circuit.

A simple Zener-diode voltage regulator (Fig. 12.18) consists of a series voltage dropping resistance R_S and a Zener diode in parallel with the load resistance R_L. Voltage V_1 is the dc output of a rectifier-filter. The function of the regulator is to keep V_L nearly constant with changes in V_1 or I_L. The operation is based on the fact that, in the Zener breakdown region, small changes in diode voltage are accompanied by large changes in diode current (see Fig. 12.6). The large currents flowing through R_S produce voltages that compensate for changes in V_1 or I_L.

Example 6

A load draws a current varying from 10 to 100 mA at a nominal voltage of 100 V. A regulator consists of $R_S = 200\ \Omega$ and a Zener diode represented by $V_Z = 100$ V and $R_Z = 20\ \Omega$ (Fig. 12.19).

For $I_L = 50$ mA, determine the allowable variation in V_1 corresponding to a 1% increase in V_L.

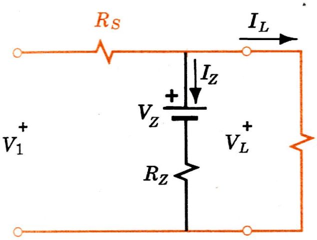


Figure 12.19 Performance of a voltage regulator.

To place the diode well into the Zener breakdown region, it is customary to use a minimum current $I_Z(\text{min}) = 0.1 I_L(\text{max}) = 0.1 \times 100 = 10$ mA. Then

$$V_L = V_Z + I_Z R_Z = 100 + 0.01(20) = 100.2\text{ V}$$

and

$$\begin{aligned} V_1 &= V_L + (I_L + I_Z)R_S \\ &= 100.2 + (0.05 + 0.01)200 = 112.2\text{ V} \end{aligned}$$

If V_L increases by 1% or 1 V, $V_L' = 101.2$ V,

$$I_Z' = \frac{V_L' - V_Z}{R_Z} = \frac{101.2 - 100}{20} = 0.06\text{ A}$$

and

$$\begin{aligned} V_1' &= V_L' + (I_L + I_Z')R_S \\ &= 101.2 + (0.05 + 0.06)200 = 123.2\text{ V} \end{aligned}$$

Conclusion: A change in V_1 of 11 V produces a change in V_L of only 1 V.

Waveshaping Circuits

A major virtue of electronic circuits is the ease, speed, and precision with which voltage and current waveforms can be controlled. Some of the basic waveshaping functions are illustrated by a radar pulse-train generator. The word *radar* stands for *ra*dio *d*etection *a*nd *r*anging. A very short burst of high-intensity radiation is transmitted in a given direction; a return echo indicates the presence, distance, direction, and speed of a reflecting object. The operation of a radar system requires a precisely formed series of timing pulses. Typically, these may be of 5 μsec duration with a repetition rate of 500 pulses per sec. Starting with a 500-Hz sinusoidal generator, the pulse train could be developed as shown in Fig. 12.20. Can you visualize some relatively simple electronic circuits that would perform the indicated functions?

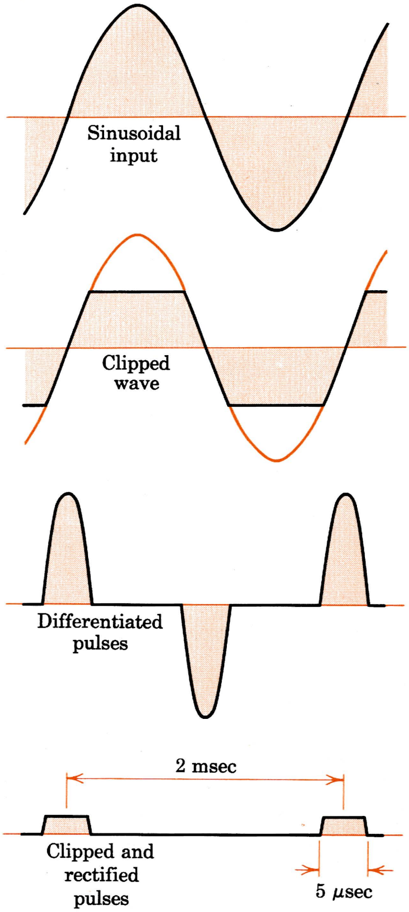


Figure 12.20 Generation of a train of timing pulses.

Differentiating

There are special amplifiers for precise differentiation and integration (described in Chapter 19), but within limits the simple circuit of Fig. 12.21 provides an output that is the derivative of the input. For the special case of a rectangular input voltage wave, the output voltage is proportional to the RC charging current in response to a step input voltage. In this case a linear circuit transforms a rectangular wave into a series of short pulses if the time constant $\tau = RC$ is small compared to the period of the input wave.

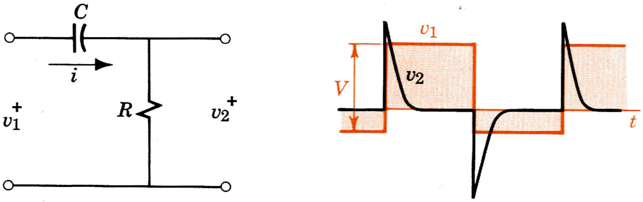
The general operation of this circuit is revealed if we make some simplifying assumptions. Applying Kirchhoff's voltage law to the left-hand loop,

$$v_1 = v_C + v_R \cong v_C \qquad (12\text{-}21)$$

if v_R is small compared to v_C. Then

$$v_2 = v_R = Ri = RC\frac{dv_C}{dt} \cong RC\frac{dv_1}{dt} \qquad (12\text{-}22)$$

The output is approximately proportional to the derivative of the input.



(a) (b)

Figure 12.21 Differentiating circuit.

Integrating

From our previous experience with circuits we expect that if differentiating is possible, integrating is also. In Fig. 12.22, a square wave of voltage has been applied long enough for a cyclic operation to be established. The time constant RC is a little greater than the half-period of the square wave. The capacitor C charges and discharges on alternate half-cycles, and the output voltage is as shown.

If the time constant RC is large compared to the period T of the square waves, only the straight portion of the exponential appears and the output is the *sawtooth* wave in which voltage is directly proportional to time. In general,

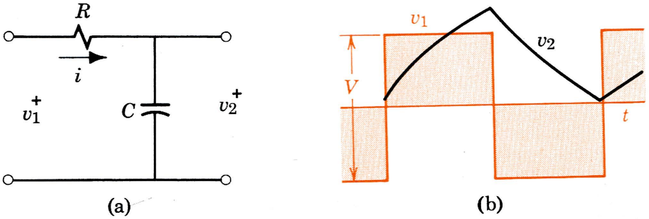


Figure 12.22 Integrating circuit.

$$v_1 = v_R + v_C \cong v_R = iR \tag{12-23}$$

if v_C is small compared to v_R (i.e., $RC > T$). Then

$$v_2 = \frac{1}{C}\int i\,dt \cong \frac{1}{RC}\int v_1\,dt \tag{12-24}$$

and the output is approximately proportional to the integral of the input. If it is necessary, the magnitude of the signal can be restored by linear amplification.

Clipping

A *clipping* circuit provides an output voltage v_2 equal to (or proportional to) the input voltage v_1 up to a certain value V; above V the wave is clipped off. If both positive and negative peaks are to be clipped, the desired transfer characteristic is as shown in Fig. 12.23a. The switching action of diodes is used

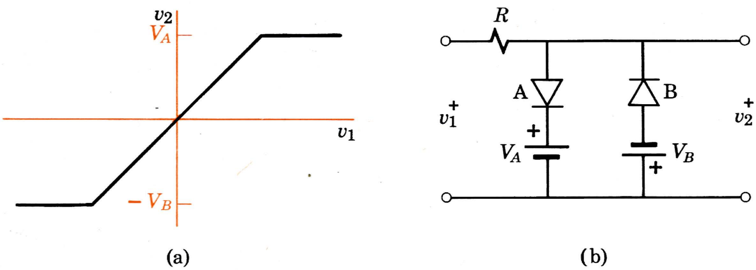


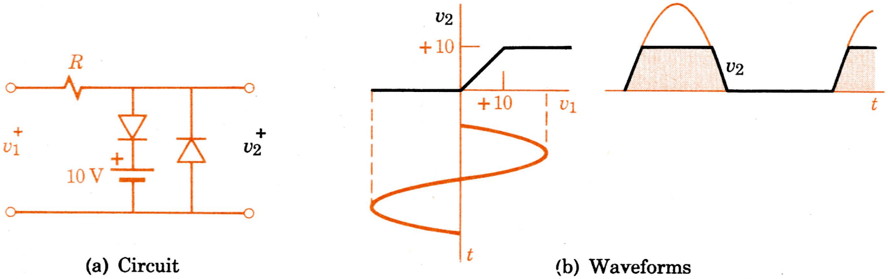
Figure 12.23 Clipping characteristic and circuit.

to provide clipping in Fig. 12.23b. The bias is set so that diode A conducts whenever $v_1 > V_A$ and diode B conducts whenever $v_1 < -V_B$. When $-V_B < v_1 < V_A$, neither diode conducts and voltage v_1 appears across the output terminals. When either diode is conducting, the difference between v_1 and v_2 appears as a voltage drop across R. The effect of an asymmetric clipping circuit on a sinewave is illustrated in the following example.

Example 7

A sinewave $v_1 = 20 \sin \omega t$ V is applied to the circuit of Fig. 12.24a. Predict the output voltage v_2.

In this circuit the first diode and the 10-V battery provide clipping for voltages greater than +10 V. With $V_B = 0$, the second diode provides clipping of all negative voltages or rectification. The circuit characteristic and resulting output are as shown in Fig. 12.24b.



(a) Circuit (b) Waveforms

Figure 12.24 A clipping and rectifying circuit.

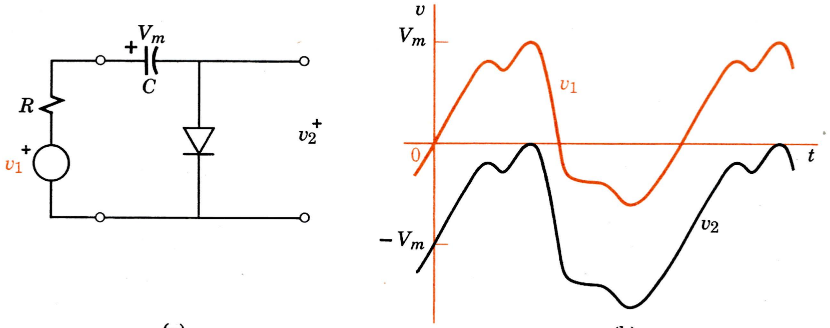
Clamping

To provide satisfactory pictures in television receivers, the peak values of certain variable signal voltages must be held or *clamped* at predetermined levels. In passing through *RC*-coupled amplifiers, the reference level is lost and a *clamper* or *dc restorer* is used.

In the circuit of Fig. 12.25a, if R is small the capacitor tends to charge up to the peak value of the input wave, just as in the half-wave rectifier with capacitor. When the polarity of v_1 reverses, the capacitor voltage remains at V_m because the diode prevents current flow in the opposite direction. Neglecting the small voltage across R, the output voltage across the diode is

$$v_2 = v_1 - V_m \tag{12-25}$$

The signal waveform is unaffected, but a dc value just equal to the peak value



(a) (b)

Figure 12.25 A diode-clamping circuit.

of the signal has been introduced. The positive peak is said to be clamped at zero.

If the amplitude of the input signal changes, the dc voltage across C also changes (after a few cycles) and the output voltage again just touches the axis. If the diode is reversed, the negative peaks are clamped at zero. If a battery is inserted in series with the diode, the reference level of the output may be maintained at voltage V_B.

Clamping and rectifying are related waveshaping functions performed by the same combination of diode and capacitor. In the rectifier the variable component is rejected and the dc value is transmitted; in the clamper the variable component is transmitted and the dc component is restored.

SUMMARY

- The important characteristic of a diode is the discrimination between forward and reverse voltages.
  The ideal diode presents zero resistance in the forward direction and infinite resistance in the reverse direction; it functions as a selective switch.
  Actual diodes depart from the ideal to some extent. Circuit models of the required precision can be derived from the characteristic curves.

- A rectifier converts alternating current into unidirectional current. In half-wave rectification with a resistive load,

$$I_{\mathrm{dc}} \cong \frac{1}{2\pi} \int_0^{\pi} \frac{V_m \sin \omega t}{R_L} d(\omega t) = \frac{1}{\pi} \frac{V_m}{R_L} = \frac{I_m}{\pi}$$

  A bridge circuit or phase inverter permits full-wave rectification.
  In full-wave rectification with a resistive load, $I_{\mathrm{dc}} = 2I_m/\pi$.

◆ The effectiveness of rectification is measured by the ripple factor

$$r = \frac{V_{\text{ac}}}{V_{\text{dc}}} = \frac{I_{\text{ac}}}{I_{\text{dc}}} = \sqrt{\left(\frac{I_{\text{rms}}}{I_{\text{dc}}}\right)^2 - 1}$$

◆ A filter reduces the ripple by absorbing voltage variations or opposing current variations or both.

A capacitor filter stores charge on voltage peaks and delivers charge during voltage valleys; an approximate analysis shows that

$$r \cong \frac{1}{2\sqrt{3}fCR_L} \qquad \text{(with half-wave rectification)}$$

In **L**-section filters, the inductor opposes current changes and the capacitor absorbs voltage peaks; ac circuit analysis shows that

$$r \cong \frac{0.47}{4\omega^2 LC} \qquad \text{(with full-wave rectification)}$$

◆ A voltage regulator minimizes changes in dc load voltage.

In Zener diodes small voltage changes cause large current changes that produce voltage drops to compensate for variations in V_{in} or I_{out}.

◆ Waveforms can be shaped easily, rapidly, and precisely.

An *RC* circuit can perform differentiation or integration.

A diode-resistor-battery circuit can perform clipping.

A peak-charging capacitor can clamp signals to desired levels.

REVIEW QUESTIONS

1. Why do we wish to replace actual devices with fictitious models?

2. Why do we insist on *linear* models?

3. Cite an example from aeronautical, chemical, civil, industrial, and mechanical engineering of a process or a device that is customarily analyzed in terms of a mathematical or physical model.

4. In what respects does a vacuum diode differ from an ideal diode? A gas diode? A semiconductor diode? A Zener diode?

5. Explain what is meant by a "linearized characteristic curve."

6. Given a *real* semiconductor diode in a half-wave rectifier circuit, sketch the voltage waveform across a load resistor. Repeat for an *ideal* diode.

7. Explain the operation of a full-wave bridge rectifier.

8. Explain the operation of a capacitor filter after a full-wave rectifier.

9. During capacitor discharge, why doesn't some of the current flow through the diode?

10. What is the effect on the ripple factor of the circuit in Fig. 12.12 of doubling R_L? Of halving C? Of doubling V_m?

11. How does the maximum allowable diode current limit the value of C?

12. Does Eq. 12-18 apply if L is too small? Why?

13. Sketch and explain the operation of a voltage regulator using the gas diode of Fig. 12.1d.

14. Explain in words the operation of a differentiator and an integrator.

15. Sketch differentiating and integrating circuits without capacitors.

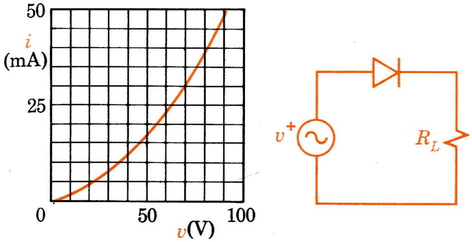
16. Explain the operation of a diode clipper.

17. A circuit that clips tops and bottoms is sometimes called a "slicer." Sketch a slicer circuit and explain its operation.

18. Sketch a clamping circuit and explain its operation.

EXERCISES

1. Represent the diode of Fig. 12.26a by a circuit model that will hold, approximately, for currents $25 < i < 50$ mA. Use your model to predict the current flowing under an applied voltage $v = 80$ V; check graphically.
2. A diode having the characteristics of Fig. 12.26a is to be used in the circuit shown.



(a) (b)

Figure 12.26 Diode characteristics.

   (a) Represent the diode (in the region $0 < v < 50$ V) by a circuit model consisting of an ideal diode and an average value of forward resistance.
   (b) Calculate the average current delivered to $R_L = 3\ \text{k}\Omega$ if $v = 50 \sin \omega t$ V.
   (c) Sketch one cycle of the applied voltage and show the current i and the voltage across the diode v_D as functions of time.
3. (a) Represent the real diode of Fig. 12.27 by a circuit model consisting of an ideal diode and a voltage source.

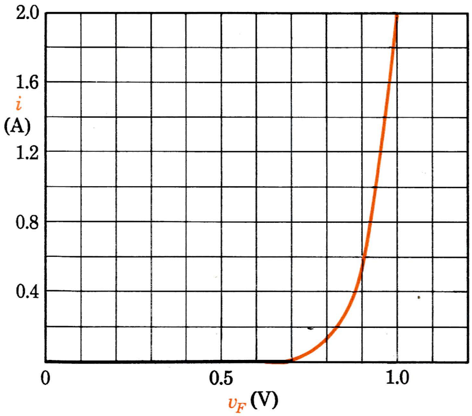


Figure 12.27 Silicon diode characteristics.

   (b) Use your model to estimate the current if the diode is placed in series with a 20-V battery and a 10-Ω resistance.
4. The diode of Fig. 12.27 is to be used in a rectifier circuit consisting of a voltage $v = 300 \cos \omega t$ V and a resistance $R_L = 200\ \Omega$. Devise an appropriately simple circuit model and predict I_{dc}.
5. Devise a circuit model to represent, approximately, the Zener diode of Fig. 12.28.

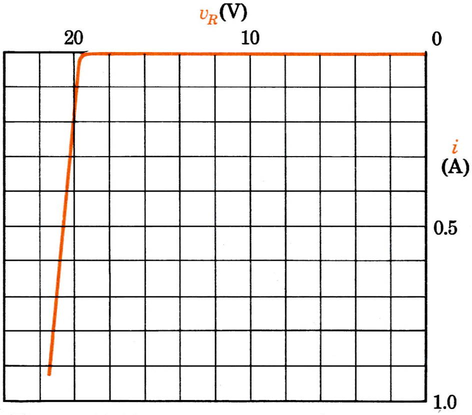


Figure 12.28 Zener region characteristics.

6. Derive approximate circuit models for the diodes characterized in Fig. 12.29.

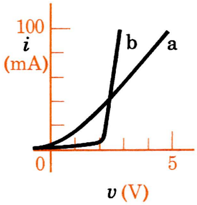


Figure 12.29

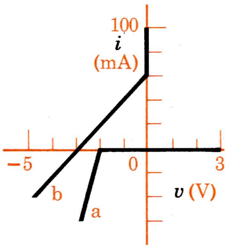


Figure 12.30

7. Derive circuit models for the devices characterized in Fig. 12.30.
8. The Zener diode whose reverse characteristics are given in Fig. 12.28 is connected

in series with a voltage $v = 40 \cos \omega t$ V and a resistance $R_L = 2\ \text{k}\Omega$. Draw on the same graph waveforms $v(\omega t)$ and $i_L(\omega t)$.

9. An ideal diode is used in a half-wave rectifier with power supplied at 120 V (rms) and 60 Hz. For a load $R_L = 2000\ \Omega$, predict I_{dc}, V_{dc}, and the power delivered to the load.

10. Repeat Exercise 9 assuming a full-wave bridge rectifier circuit.

11. A vacuum diode with a forward resistance $R_F = 100\ \Omega$ is connected in series with a voltage $v = 100 \cos \omega t$ V and $R_L = 400\ \Omega$.
   (a) On the same graph, sketch v, i, and v_D as functions of time.
   (b) Predict the readings on a d'Arsonval type dc ammeter and an electrodynamometer-type ac ammeter connected in series with R_L.
   (c) Predict the readings on dc voltmeters connected across R_L and across the diode. Does $I_{dc}R_F = V_{dc}$?

12. The ratio of dc power delivered to the load to power supplied to the rectifier circuit (rectifier and load) is *rectification efficiency*. Assuming $R_L = 10\ \text{k}\Omega$, $R_F = 2\ \text{k}\Omega$ for the diode, and a sinewave voltage input:
   (a) Calculate the rectification efficiency of a half-wave circuit.
   (b) Repeat part (a) for a full-wave bridge rectifier using four identical diodes.
   (c) Repeat part (a) for a full-wave rectifier using a center-tapped transformer.

13. The diode of Fig. 12.27 is connected in series with a 30-V rms source to charge a 12-V battery with an internal resistance of 0.1 Ω. Specify the series resistance necessary to limit the peak current to 2 A. Estimate the time required to recharge a 10-A-hr battery.

14. Calculate the ripple factor for a rectangular wave of amplitude 20 mA that is "on" for 3 ms and "off" for 1 ms.

15. A half-wave rectifier with capacitor filter (Fig. 12.12a) supplies 200 V dc to R_L. For $C = 60\ \mu\text{F}$ and $R_L = 4\ \text{k}\Omega$, $V_m = 208$ V, $\theta = 22°$, and $I_m = 1.7$ A. Predict the new values of V_m, θ, and I_m for:
   (a) $C = 60\ \mu\text{F}$ and $R_L = 8\ \text{k}\Omega$;
   (b) $C = 30\ \mu\text{F}$ and $R_L = 4\ \text{k}\Omega$.

16. A 60-Hz half-wave rectifier with a capacitor filter ($C = 40\ \mu\text{F}$) supplies 20 mA of direct current to a load resistance of 15 kΩ. Estimate the transformer secondary voltage (rms), the ripple factor, the conduction angle, and the peak diode current.

17. Repeat Exercise 16 for a full-wave rectifier.

18. Design a 60-Hz full-wave rectifier circuit with capacitor filter to supply 200 V at 50 mA with less than 1% ripple factor.
   (a) Draw a wiring diagram and sketch a graph showing supply voltage, load voltage, and diode current as functions of time.
   (b) Specify the necessary transformer secondary voltage (rms).
   (c) Using the approximate analysis; estimate the capacitance required.
   (d) Estimate the peak diode current expected.
   (e) On the graph of part (a) indicate the PIV and estimate its value.

19. Filters A and B (Fig. 12.31) are both designed to provide 2% ripple factors when operated with 60-Hz rectifiers. Predict the new ripple factors if the input frequency is increased to 120 Hz, all other factors unchanged.

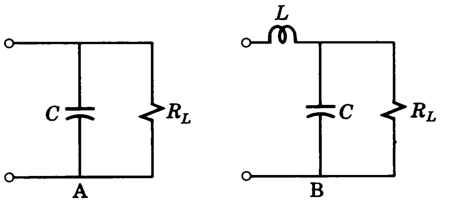


Figure 12.31

20. The output of a 60-Hz full-wave rectifier is applied to filter B in Fig. 12.31. At 60 Hz (NOTE!) the filter reactances are $X_L = 2000\ \Omega$ and $X_C = 200\ \Omega$. The dc voltage is 300 V across $R_L = 2\ \text{k}\Omega$.
   (a) Predict the ac ripple voltage (rms) in the output.
   (b) Predict the rms value of the next higher harmonic in the output.

21. The specifications of Exercise 18 are to be met by a full-wave rectifier and an L-section filter with $C = 10\ \mu\text{F}$. Assuming that the rectifier output voltage is represented by the first two terms of the Fourier series:
   (a) Calculate the required transformer

secondary voltage (rms), stating any necessary assumptions.

(b) Specify the required value of inductance L. (Is your approach valid?)

(c) Determine the factor by which the filter has reduced the magnitude of the ac voltage component.

(d) Predict, without detailed calculation, the ac voltage across R_L if two L-sections in cascade are used.

22. In Fig. 12.32a, a dc voltage source is represented by $V_S = 100$ V and $R_S = 100\ \Omega$.

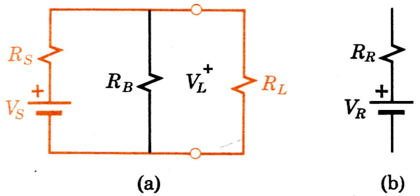


Figure 12.32

(a) Disregard R_B and calculate the percent variation in load voltage V_L as R_L varies from 1000 to 2000 Ω.

(b) Repeat with "bleeder" resistance $R_B = 100\ \Omega$ connected.

(c) Repeat with R_B replace by $V_R = 90$ V and $R_R = 10\ \Omega$

(d) Summarize the voltage regulation properties of these circuits.

(e) What passive electronic device could replace the regulator of circuit (c)?

23. In Fig. 12.33, an 8-V Zener diode ($R_Z \cong 0$) holds the load voltage at 8 V. The supply voltage V varies from 9 to 12 V and the load current I_L varies from 0 to 100 mA.

(a) Define and specify the "best" value of R_S.

(b) Predict the maximum current expected in the Zener diode.

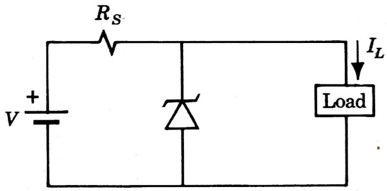


Figure 12.33

24. A voltage source with an internal resistance $R_S = 100\ \Omega$ is to supply $I_L = 100$ mA to $R_L = 310\ \Omega$. To reduce load voltage variations, a Zener diode ($V_Z = 30$ V, $R_Z = 10\ \Omega$, and I_Z(max) $= 300$ mA) regulator is to be used.

(a) Draw an appropriate schematic diagram and calculate V_S and V_L.

(b) For a 10% drop in source voltage, predict the percent change in V_L.

(c) For V_S at the value calculated in part (a) and I_L increased by 20%, predict the new V_L and the percent change in V_L.

(d) For $R_L = 310\ \Omega$, if V_S drops to 29 V, predict the new I_Z and V_L.

25. What function is performed by the circuit of Fig. 12.34a? Stating any necessary assumptions, support your conclusion mathematically.

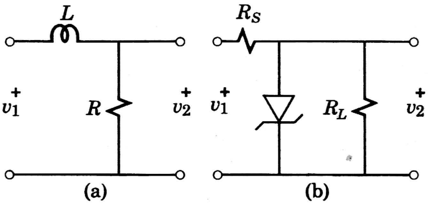


Figure 12.34

26. What function is performed by the Zener diode circuit of Fig. 12.34b? For $v_1 = 20 \sin \omega t$ V, $V_Z = 5$ V, and $R_L \gg R_S$, plot $v_2(t)$.

27. In Fig. 12.23b, $R = 1$ kΩ, $V_A = 2$ V, and $V_B = 5$ V. Sketch v_2 for $v_1 = 6 \sin \omega t$ V.

28. For $v_1 = 10 \sin \omega t$ V, $R = 1$ kΩ, and $V = 4$ V in Fig. 12.35a, sketch $v_2(t)$.

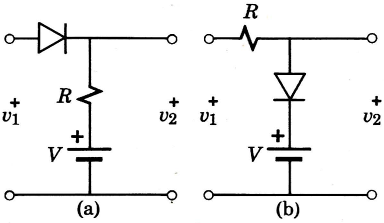


Figure 12.35

29. For $v_1 = 10 \sin \omega t$ V, $R = 1$ kΩ, and $V = 4$ V in Fig. 12.35b, sketch $v_2(t)$.

30. Draw the transfer characteristics v_2 versus v_1 for Figs. 12.35a and b.

31. A voltage v_1 varies from -3 to $+6$ V. For a certain purpose, this voltage must be limited to a maximum value of $+4.0$ V. A silicon diode, a 6-V battery, and assorted resistances are available. Design a suitable circuit.

32. In Fig. 12.36a, $v_1 = 10 \sin \omega t$ V, R is "small," and $V_B = 5$ V. Plot v_1, v_C, and v_2 on the same time axis.

33. Repeat Exercise 32 for Fig. 12.36b.

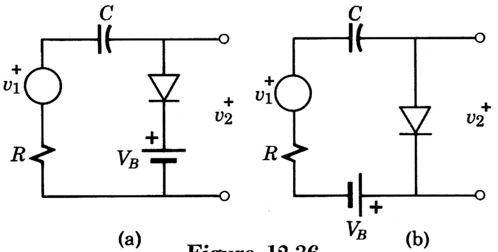


Figure 12.36

PROBLEMS

1. In Fig. 12.37, R_L is replaced by a high-resistance dc voltmeter. When a voltage $v_1 = V_m\ (0.5 + \sin \omega t)$ V is applied, the voltmeter reads 70 V. Determine V_m and define the function of this instrument.

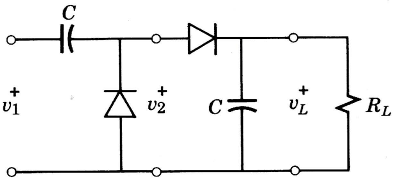


Figure 12.37

2. In the gating circuit of Fig. 12.38, input voltage $v_1 = 10 \sin 2\pi 100t$ V and switching voltage v_G is a square wave of amplitude ± 10 V and frequency 100 Hz. Draw several cycles of v_1, v_G, and v_2 on the same time axis.

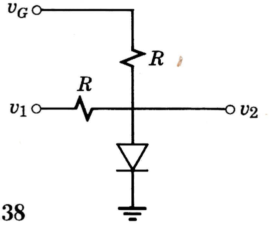


Figure 12.38

3. Given a current $i(t) = I_{\text{dc}} + i_{\text{ac}}$ where I_{dc} is the average value and i_{ac} is the varying component of a periodic wave, derive Eq. 12-5.

4. A technician constructing the filter of Example 5 suggests replacing inductor L with "a much cheaper resistor with the same impedance." If this were done, what would be the ripple factor and the PIV? Is this a practical solution? Under what circumstances would this be a practical solution?

5. The Zener diode of Fig. 12.6b is used in the voltage regulator circuit of Fig. 12.39, where $R_S = 1000\ \Omega$ and $R_L = 2000\ \Omega$. If V_1 increases from 16 to 24 V (a 50% increase), calculate the corresponding variation in load voltage V_L. Is the "regulator" doing its job?

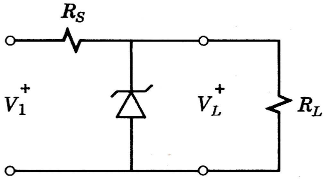


Figure 12.39

6. A precise wave shaper can be built using an amplifier as shown in Fig. 12.40. Since the amplifier gain is very large, the input voltage $v_i = v_2/(-A) \cong 0$. Since R_i is very large compared to R_1, the input current $i_i \cong 0$. Using these, derive an expression for v_2 in terms of v_1 and identify the function performed.

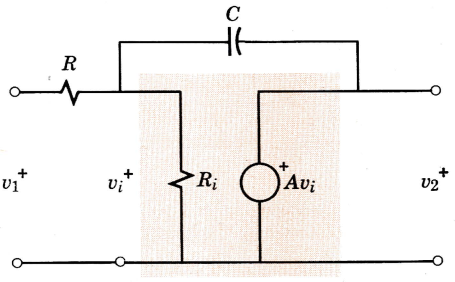


Figure 12.40

Chapter 13

Transistors and Integrated Circuits

Field-Effect Transistors
Junction Transistors
Integrated Circuits
Thyristors

In the idealized transducers described on page 274, the output current is controlled by the input current or the input voltage. In either case, amplification is possible by sending the output current through a load resistance. We have seen how such amplification is achieved with vacuum triodes and pentodes. The long search for a semiconductor device capable of performing the same functions, but without the disadvantages of the thermionic tube, led to the discovery of the *transistor.*

The transistor is available in several basic forms for a wide variety of applications ranging from implantable heart pacers to giant digital computers. In its most common configuration, it may be a tiny portion of a complex integrated circuit or it may serve as a discrete switch controlling many amperes.

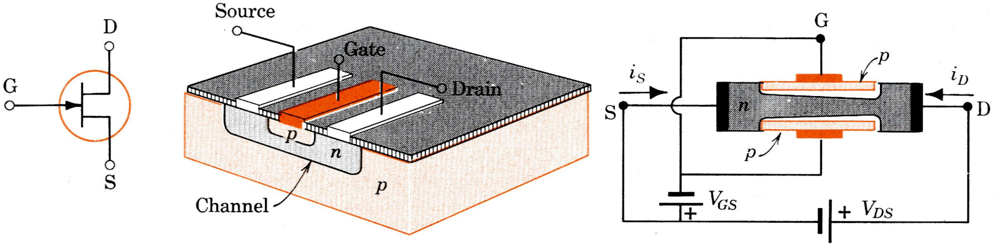
In this chapter we develop the principles necessary for the analysis and design of circuits containing transistors. We start with the simplest form of "unipolar" transistor and derive the *i-v* characteristics that determine its behavior. Then we study the more complicated "bipolar" transistor and derive its characteristics. Next we see how combinations of transistors, diodes, resistors, and capacitors can be fabricated into "integrated circuits." Finally we learn how multijunction thyristors can be used to control large amounts of power efficiently.

Field-Effect Transistors

The control of anode current in a vacuum tube by varying the electric field adjacent to the cathode has its semiconductor counterpart in the *field-effect transistor.* Shockley proposed the device in 1952, but only recently have the fabrication techniques necessary for dependable operation become available.

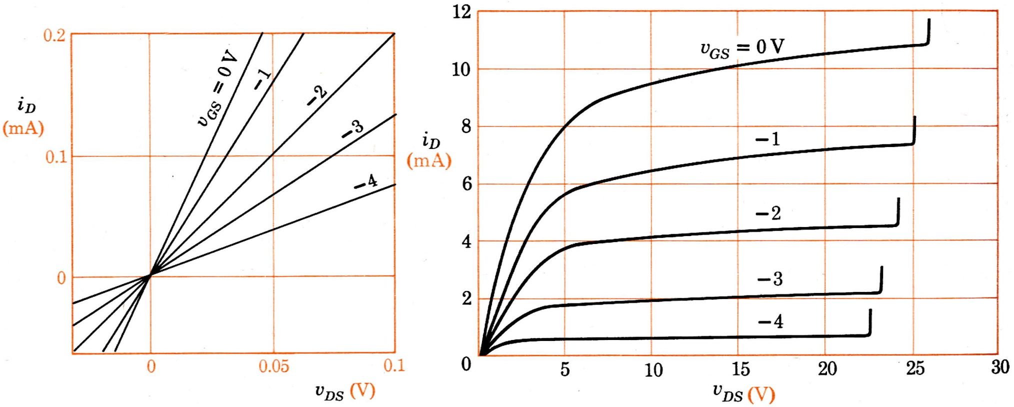
JFET

In the *junction-gate* form of FET (Fig. 13.1b), a thin conducting *channel* of finite conductance is established between two *pn* junctions. (The arrow in the symbol always points toward *n*-type material.) The current from the *source* to the *drain* for a given voltage v_{DS} depends on the dimensions of the channel. If the *pn* junctions are reverse biased by applying a voltage to the *gate*, a depletion region containing no mobile carriers is formed and the width of the conducting channel is reduced. As shown by the low-voltage characteristics (Fig. 13.2a), the conductance $\Delta i_D/\Delta v_{DS}$ is dramatically affected by the gate-source voltage v_{GS}. The depletion region is asymmetric because the reverse bias is highest at the drain end of the channel where the difference in potential between gate



(a) Symbol (b) Typical structure (c) Simplified representation

Figure 13.1 An *n*-channel depletion-mode field-effect transistor (JFET).



(a) Low voltage (b) Normal voltage

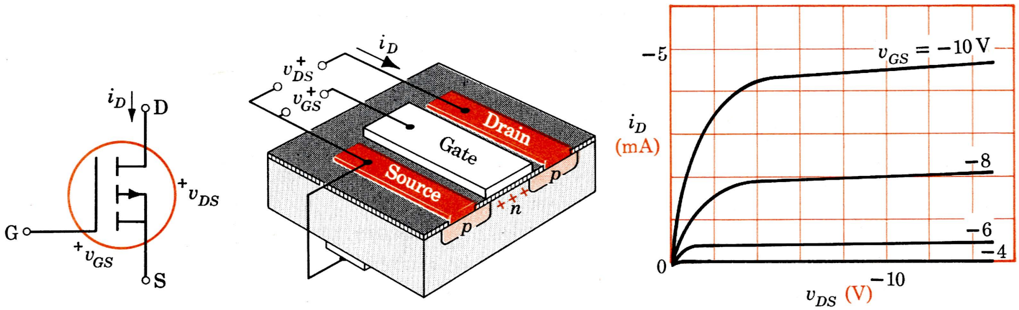
Figure 13.2 Depletion-mode field-effect transistor characteristics.

and channel is highest. For larger values of v_{DS}, the reverse bias increases until the two depletion regions almost meet, tending to "pinch off" the conducting channel. In Fig. 13.2b the *pinch-off voltage* V_p for $v_{GS} = 0$ is about 5 V. Above pinch-off, an increase in v_{DS} decreases the channel width, offsetting an increase in current density, and the i_D curve flattens out.

If a negative voltage is applied to the gate, pinch-off occurs at a lower drain-source voltage and the drain current is limited to a lower value. Above pinch-off, the current curves are relatively flat until the gate-drain voltage reaches the point where avalanche *breakdown* occurs. The normal operating range extends from pinch-off to breakdown.

MOSFET

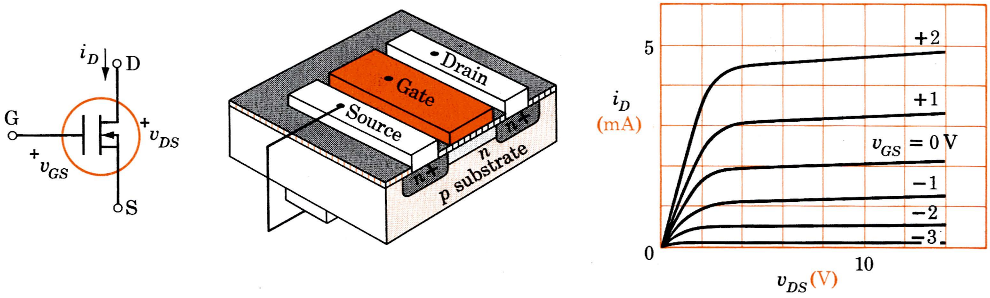
In the metal-oxide-semiconductor FET or MOSFET, a thin layer of SiO_2 insulates the gate contact from the channel. Of greatest commercial importance is the *p-channel enhancement-mode* transistor (Fig. 13.3). In this device, a conducting channel is induced by an electric field established between the gate and the n-type substrate. With no gate voltage, little current flows through the two back-to-back pn junctions. With a small negative gate voltage, electrons in the adjacent n material are repelled and a depletion layer formed. At a slightly more negative voltage, an *inversion layer* of mobile holes is formed at the surface of the n-type region, which becomes p-type. The conductivity of the region has been "enhanced," the transistor has been "turned on," and current can flow between source and drain. The current is not proportional to v_{DS}, however. As v_{DS} becomes more negative, the gate-to-channel voltage and the accompanying electric field are reduced, and the current levels off.



(a) Symbol (b) Typical structure (c) Characteristics

Figure 13.3 A p-channel enhancement-mode MOSFET.

In another form of MOSFET, a narrow lightly doped conducting layer is diffused into the channel region (Fig. 13.4b). At $v_{GS} = 0$, an appreciable drain current flows. Application of the appropriate gate voltage can cause depletion *or* enhancement of the conducting channel. The characteristic curves are similar



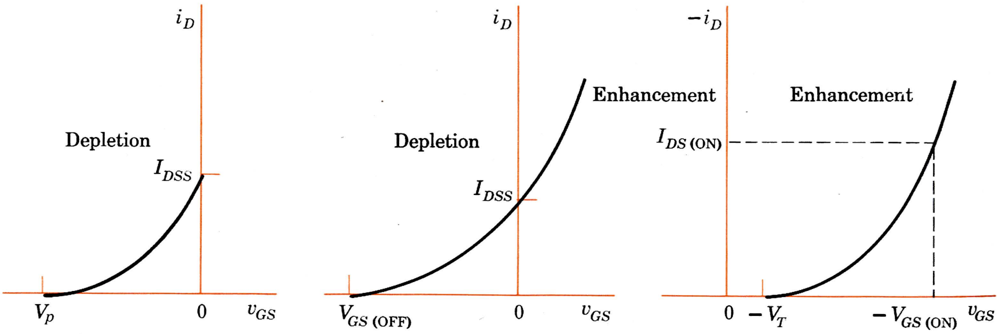
(a) Symbol (b) Typical structure (c) Characteristics

Figure 13.4 An n-channel depletion- or enhancement-mode MOSFET.

to those for the JFET with the added flexibility of permitting positive or negative control voltages; however, the practical n-channel device is more difficult to produce commercially.

Transfer Characteristics

The *i-v* characteristics of field-effect transistors resemble those of the idealized transducer (Fig. 10.21b) in which output current is controlled by input voltage, and the FET can be used as an amplifier or as a switch. Since the characteristics of individual devices are never precisely known, approximate methods of analysis are acceptable. Within the normal operating region, i.e., between pinch-off or turn-on and breakdown, drain current i_D is nearly independent of drain-source voltage v_{DS}, and the *transfer characteristics* are as shown in Fig. 13.5.



(a) JFET (b) DE MOSFET (c) E MOSFET

Figure 13.5 Transfer characteristics for three types of FET.

From theoretical analysis and practical measurement it can be shown that the transfer characteristics are approximately parabolic in all three cases. For the JFET, the drain current in the constant-current region is

$$i_{DS} = I_{DSS}(1 - v_{GS}/V_p)^2 \tag{13-1}$$

where i_{DS} = the drain current in the constant-current region,
I_{DSS} = the value of i_{DS} with gate shorted to source, and
V_p = the pinch-off voltage.

The depletion-*or*-enhancement MOSFET is also described by Eq. 13-1, and both positive and negative values of v_{GS} are permitted. For the enhancement-only MOSFET, the corresponding relation is

$$i_{DS} = K(v_{GS} - V_T)^2 \tag{13-2}$$

where K is a constant of proportionality, and V_T is the turn-on or *threshold voltage.*

These simple relations are useful in predicting the dc behavior of FETs. In practice, the manufacturer usually specifies typical values of I_{DSS} and the *gate-source cutoff voltage* $V_{GS(OFF)}$, which is approximately equal to V_p since the same pinch-off effect is created between gate and channel. For the enhancement-only MOSFET, the manufacturer specifies V_T and a particular value of $I_{DS(ON)}$ corresponding to a specified value of $V_{GS(ON)}$.

Example 1

For the enhancement only MOSFET of Fig. 13.6, the manufacturer specifies $V_T = -4$ V and $I_{DS} = -7.2$ mA at $V_{GS} = -10$ V. For $V_{DD} = 24$ V and $R_G = 100$ MΩ, specify R_D for operation at $V_{DS} = -8$ V.

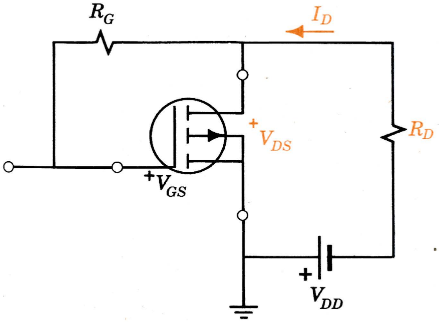


Figure 13.6 MOSFET analysis.

Substituting the given data in Eq. 13-2,

$$I_{DS} = -0.0072 = K(V_{GS} - V_T)^2 = K(-10 + 4)^2$$
$$\therefore K = -0.0072/(6)^2 = -0.0002$$

Since $i_G = 0$, there is no drop across R_G; therefore, $V_{GS} = V_{DS} = -8$ V, and

$$I_D = -0.0002(-8 + 4)^2 = -3.2 \text{ mA}$$

$$R_D = \frac{V_{DD} - V_{DS}}{I_D} = \frac{-24 + 8}{-0.0032} = 5 \text{ k}\Omega$$

Because the gate-source junction of the JFET is reverse biased, the input signal current is very small; in other words, the input resistance is very high and little input power is required. In the MOSFET or *insulated-gate* FET, the input resistance may be as high as 10^{15} Ω.† Disadvantages of the FET include its higher operating voltage, lower voltage gain, and poorer high-frequency performance in comparison to the more common type of transistor.

Example 2

The MOSFET of Fig. 13.4 is to be used as an amplifier. Assuming $I_{DSS} = 2$ mA and $V_{GS(\text{OFF})} \cong -4$ V, predict the voltage gain with a load resistance $R_L = 4$ kΩ.

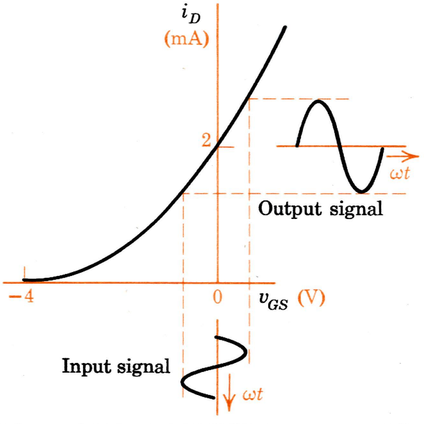


Figure 13.7 A MOSFET as an amplifier.

The transfer characteristic (Eq. 13-1) becomes

$$i_D = 0.002(1 + v_{GS}/4)^2$$

For a signal $v_{GS}(t) = V_m \sin \omega t$ (Fig. 13.7), the output voltage across $R_L = 4$ kΩ will be

$$\begin{aligned} v_L = i_D R_L &= 8(1 + 0.25V_m \sin \omega t)^2 \\ &= 8 + 4V_m \sin \omega t + 0.5V_m^2 \sin^2 \omega t \end{aligned}$$

The output consists of an 8-V dc component, a $4V_m$ signal component at the input frequency, and a $0.5V_m^2$ "distortion" component resulting from the nonlinear (parabolic) transfer characteristic.

For small values of V_m, this is an amplifier with a voltage gain of 4.

For large values of V_m, this is a "square-law device" with special virtues described in Chapter 17.

Junction Transistors

The most widely employed electronic device is the *bipolar junction transistor.*‡ In contrast to the "unipolar" FET, in the "bipolar" device both majority and minority carriers play significant roles. In comparison to the FET, the bipolar junction transistor, to be called "transistor" henceforth, permits much greater gain and provides better high-frequency performance. Essentially, the transistor consists of two *pn* junctions in close proximity. Like the diode, it is formed of a single crystal with the doping inpurities distributed so as to create two abrupt changes in carrier density.

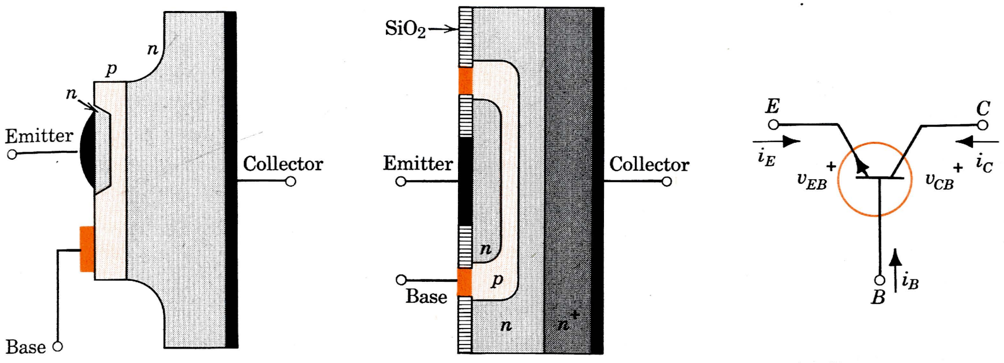
---

† The nearly infinite gate resistance creates a practical handling problem. If a small static charge (10^{-10} C) builds up on the gate-substrate capacitance (10^{-12} F), the voltage produced ($V = Q/C = 100$ V) may rupture the gate. External gate leads from an *IC* are usually protected by a Zener diode.

‡ Semiconductor experiments at Bell Telephone Laboratories led to new theoretical concepts, and William Shockley proposed an idea for a semiconductor amplifier that would critically test the theory. The actual device had far less amplification than predicted, and John Bardeen suggested a revision of the theory. In December 1947, Bardeen and Walter Brattain discovered a new phenomenon and created a novel device—the *point-contact transistor.* The next year Shockley invented the *junction transistor.*

Fabrication

The structure of an *npn* alloy-diffused transistor is shown in Fig. 13.8a. An *n*-type semiconductor chip less than 1 mm square serves as the *collector* and provides mechanical strength for mounting. A *p*-type *base* region is created by diffusion and a connection is provided by a metallic contact. An *n*-type *emitter* region is then alloyed to the base region. The result is a pair of *pn* junctions separated by a base region that is thinner than this paper.

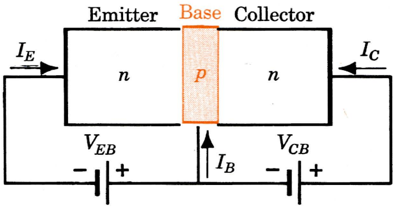


(a) Alloy structure (b) Planar structure (c) Circuit symbol

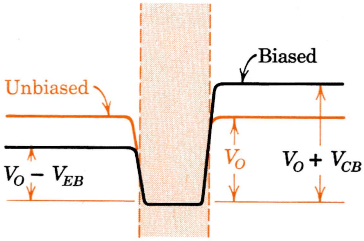
Figure 13.8 Types of *npn* transistors.

In the planar structure (Fig. 13.8b), a lightly doped film (n) is grown epitaxially upon a heavily doped (n^+) substrate. After oxidation of the surface, a window is opened by etching and an impurity (p) is allowed to diffuse into the crystal to form a junction. After reoxidation, a smaller window is opened to permit diffusion of the emitter region (n). Finally, contacts are deposited and leads attached.

In the conventional symbol (Fig. 13.8c), the emitter lead is identified by the arrow that points in the direction of positive charge flow in normal operation. (A *pnp* transistor is equally effective. The arrow in the symbol always points toward *n*-type material.) Although an *npn* transistor will work with either *n*-region serving as the emitter, it should be connected as labeled because doping densities and geometries are intentionally asymmetric.



(a) Biasing conditions



(b) Potential distribution

Figure 13.9 Operation of an *npn* transistor.

Operation

Figure 13.9a represents a thin horizontal slice through an *npn* transistor. The operation can be explained qualitatively in terms of the potential distributions across the junctions (Fig. 13.9b). The emitter junction is forward biased; the effect of bias voltage V_{EB} is to reduce the potential barrier at the emitter junction and to facilitate the injection of electrons (in this example) into the base where they are minority carriers. The collector junction is reverse-biased; the effect of bias voltage V_{CB} is to increase the potential barrier at the collector junction. The base is so thin that almost all the electrons injected into the base from the emitter diffuse across the base and are swept up the potential hill into the collector where they recombine with holes "supplied" by the external battery. (Actually, electrons are "removed" by the external battery, leaving a supply of holes available for recombination.) The net result is the transfer from the emitter circuit to the collector circuit of a current that is nearly independent of the collector-base voltage. As we shall see, this transfer permits the insertion of a large load resistance in the collector circuit to obtain voltage amplification. The modulation of the emitter-junction potential barrier by a small-signal voltage produces a corresponding variation in emitter and collector currents and a relatively large output voltage across the load resistance. Alternatively, variation of the base current (the small difference between emitter and collector currents) can be used to control the relatively larger collector current to achieve current amplification.

DC Behavior

The dc behavior of a transistor can be predicted on the basis of the motion of charge carriers across the junctions and into the base. With the emitter junction forward biased and the collector junction reverse biased, in so-called *normal* operation, the carrier motion of an *npn* transistor is as shown in the symbolic diagram of Fig. 13.10. The emitter current i_E consists of electrons injected across the *np* junction and holes injected from the base. To optimize transistor performance, the base is doped relatively lightly, emitter efficiency γ is almost unity, and most of the current consists of electrons injected from the emitter. Some of the injected electrons recombine with holes in the *p*-type base, but the base is made very narrow so that most of the electrons (minority carriers in this *p*-region) diffuse across the base and are swept across the collector junction (up the potential hill). The factor α varies from 0.90 to 0.999; a typical value is 0.98.

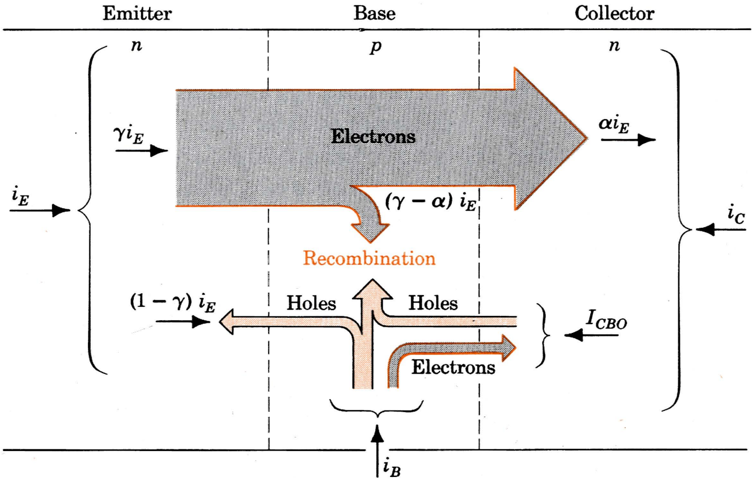


Figure 13.10 Carrier motion in the normal operation of an *npn* transistor.

The electron current αi_E constitutes the major part of the collector current. In addition, there is a reverse saturation current across the collector junction due to thermally generated minority carriers, just as in a diode (see Eq. 11-21). In a transistor, this *collector cutoff current* is given by

$$-I_{CBO}(e^{eV/kT} - 1) \cong I_{CBO} \tag{13-3}$$

when the reverse biasing exceeds a few tenths of a volt. The total collector current is then

$$i_C = -\alpha i_E + I_{CBO} \tag{13-4}$$

where α, the *forward current-transfer ratio,* increases slightly with increased collector biasing voltage v_{CB}.

The base current (Fig. 13.10) consists of holes diffusing to the emitter, positive charges to supply the recombination that occurs in the base, and the reverse saturation current to the collector. By Kirchhoff's law,

$$i_B = -i_E - i_C \tag{13-5}$$

and i_B is the small difference between two nearly equal currents.

Common-Base Characteristics

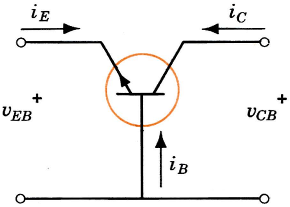
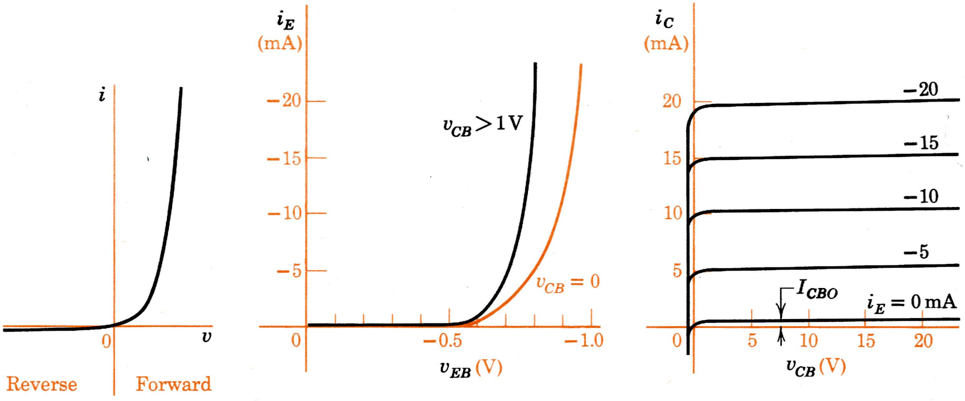


Figure 13.11 A transistor in the common-base configuration.

The transistor connection of Fig. 13.11 is called the *common-base* configuration because the base is common to input and output ports. The v-i characteristics of a transistor in this configuration can be derived from our knowledge of diode characteristics and transistor operation. Since the emitter-base section is essentially a forward-biased diode, the input characteristics in Fig. 13.12b are similar to those of the first quadrant of Fig. 13.12a; the effect of collector-base voltage v_{CB} is small. With the emitter open-circuited, $i_E = 0$ and the base-collector section is essentially a reverse-biased junction. For $i_E = 0$, $i_C \cong I_{CBO}$ (exaggerated in Figure 13.12c) and the collector characteristic is similar to the third quadrant of Fig. 13.12a. For $i_E = -5$ mA, the collector current is increased by an amount $-\alpha i_E \cong +5$ mA (see Eq. 13-4) and the curve is as shown. The slope of the curves in Fig. 13.12c is due to an effective increase in α as v_{CB} increases. Note the similarity of the common-base characteristics to those of the idealized current amplifier in Fig. 10.20. The common-base configuration is not good for practical current amplification, however, because the factor α is always less than 1.



(a) Diode characteristic **(b) Emitter characteristics** **(c) Collector characteristics**

Figure 13.12 Common base characteristics of an *npn* transistor.

Common-Emitter Characteristics

If the *npn* transistor is reconnected in the *common-emitter* (CE) configuration of Fig. 13.13, current amplification is possible. The input current is now base current i_B, and emitter current $i_E = -(i_C + i_B)$; therefore, collector current is

$$i_C = -\alpha i_E + I_{CBO} = +\alpha(i_C + i_B) + I_{CBO}$$

Solving,

$$i_C = \frac{\alpha}{1-\alpha} i_B + \frac{I_{CBO}}{1-\alpha} \qquad (13\text{-}6)$$

To simplify Eq. 13-6, we define the *current transfer ratio* (*CE*) as

$$\beta = \frac{\alpha}{1-\alpha} \qquad (13\text{-}7)$$

and note that the *collector cutoff current* (*CE*) is

$$\frac{I_{CBO}}{1-\alpha} = (1+\beta)I_{CBO} = I_{CEO} \qquad (13\text{-}8)$$

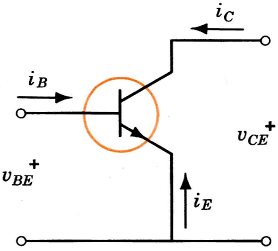


Figure 13.13 A transistor in the common-emitter configuration.

The simplified equation for output (collector) current in terms of input (base) current and the current transfer ratio (beta) is

$$i_C = \beta i_B + I_{CEO} \qquad (13\text{-}9)$$

Example 3

A silicon *npn* transistor with $\alpha = 0.99$ and $I_{CBO} = 10^{-11}$ A is connected as shown in Fig. 13.14. Predict i_C, i_E, and v_{CE}. (*Note:* It is convenient and customary in drawing electronic circuits to omit the battery, which is assumed to be connected between the +10-V terminal and ground.)

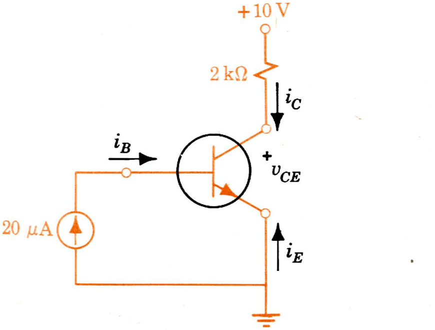


Figure 13.14 Transistor operation.

For this transistor,

$$\beta = \frac{\alpha}{1-\alpha} = \frac{0.99}{1-0.99} = \frac{0.99}{0.01} = 99$$

and the collector cutoff current is

$$I_{CEO} = (1+\beta)I_{CBO} = (1+99)10^{-11} = 10^{-9}\text{ A}$$

The collector current is

$$i_C = \beta i_B + I_{CEO} = 99 \times 2 \times 10^{-5} + 10^{-9} \cong 1.98\text{ mA}$$

As expected for a silicon transistor, I_{CEO} is a very small part of i_C.

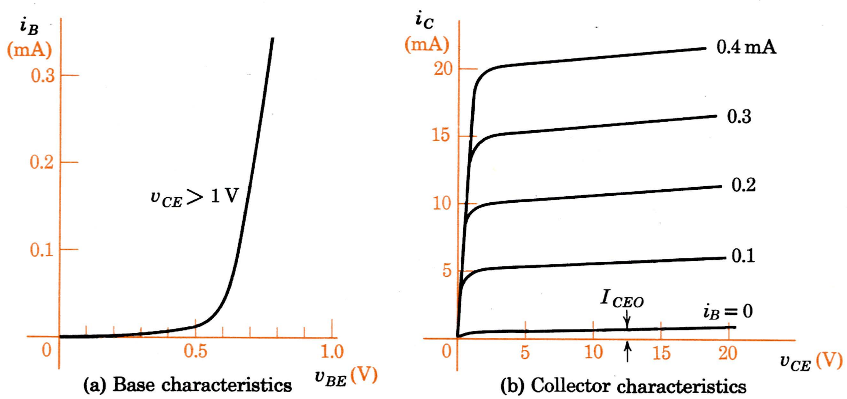
The emitter current is

$$i_E = -(i_B + i_C) = -(0.02 + 1.98)10^{-3} = -2\text{ mA}$$

The collector-emitter voltage is

$$v_{CE} = 10 - i_C R_C \cong 10 - 2(\text{mA}) \times 2(\text{k}\Omega) = 6\text{ V}$$

Since $v_{CB} = v_{CE} - v_{BE} \cong 6 - 0.7 = +5.3$ V, the collector-base junction is reverse biased.



(a) Base characteristics (b) Collector characteristics

Figure 13.15 Common-emitter characteristics of an *npn* transistor.

Typical common-emitter characteristics are shown in Fig. 13.15. The input current i_B is small and, for a collector-emitter voltage of more than a volt or so, depends only on the emitter-base junction voltage. The collector characteristics are in accordance with Eq. 13-9; for $i_B = 0$, the collector current is small and nearly constant at a value I_{CEO} (exaggerated in the drawing). For each increment of current i_B, the collector current is increased an amount βi_B. For $\alpha = 0.98$, $\beta = \alpha/(1-\alpha) = 0.98/(1-0.98) = 49$, and a small increase in i_B corresponds to a large increase in i_C. A small increase in α produces a much greater change in β, and the effect of v_{CE} is more pronounced. Note that $v_{CE} \cong v_{CB}$ since v_{BE} is so small.

Example 4

The transistor of Fig. 13.15 ($\beta \cong 50$ and $I_{CEO} \cong 1$ nA), is used in the *switching circuit* of Fig. 13.16. Predict the output voltage for input voltages of 0.2 and 0.8 V.

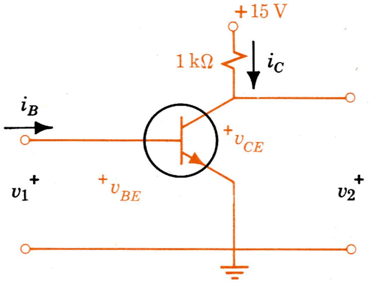


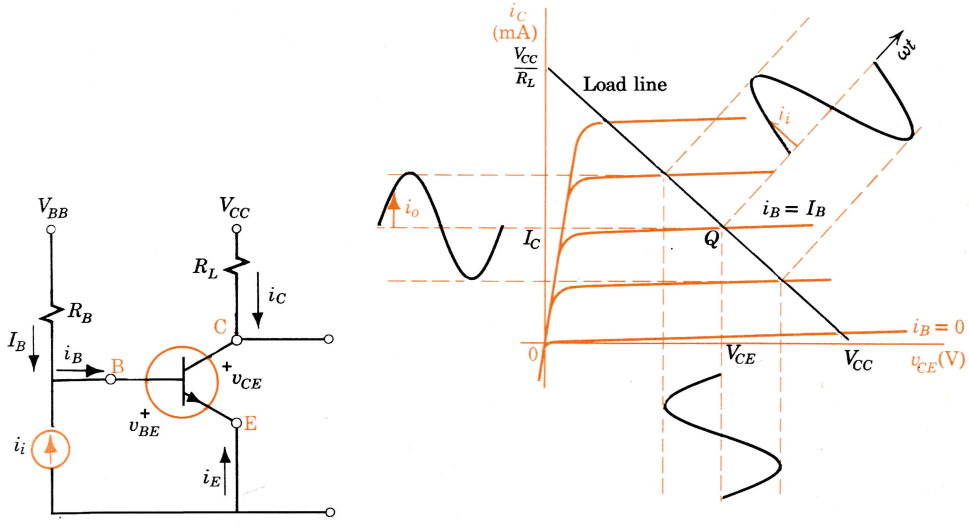
Figure 13.16 A transistor switch.

For $v_1 = v_{BE} = 0.2$ V, base current $i_B \cong 0$ and $i_C = I_{CEO} \cong 1$ nA. The voltage drop across the 1-kΩ resistor is only 1 μV and voltage $v_2 \cong 15$ V.

For $v_1 = v_{BE} = 0.8$ V, $i_B \cong 0.4$ mA. Assuming $\beta = 50$, $i_C = \beta i_B = 50 \times 0.4 = 20$ mA. But this is impossible since only 15 V is available to drive collector current through the 1-kΩ resistor.

Fig. 13.15b indicates the β concept is not valid for small values of v_{CE}; as i_B increases, v_{CE} approaches as a limit a value of less than 1 V.

Conclusion: A change in input from 0.2 to 0.8 V switches the output from 15 to 1 V.



(a) Circuit diagram (b) Operation current

Figure 13.17 An elementary *npn* transistor current amplifier.

Current Amplification

Comparison of the collector characteristics of a transistor with those of the idealized current amplifier indicates the possibility of current amplification. The graphical analysis of a transistor amplifier is similar to that of a vacuum-tube amplifier. In Fig. 13.17, batteries supply power at voltages V_{BB} and V_{CC} to the *npn* transistor whose collector characteristics are shown. The input signal (assumed sinusoidal) is applied in parallel with the emitter biasing circuit. The reverse bias on the collector is maintained by V_{CC}; note that most of this voltage appears across the collector junction since the voltage across the forward-biased emitter junction is quite small. (Why? See Fig. 13.15a.) We anticipate that a signal current i_i will cause a variation in base current i_B which will produce a variation in collector current i_C; the varying component of i_C constitutes an amplified output current.

For linear amplification, the quiescent point Q should be in the linear region on the characteristic curves. In this elementary application, let us assume that the characteristics are known precisely and that the Q point has been selected. For no signal input, $i_i = 0$, and $i_B = I_B$, the quiescent value. (See Table 13-1.) Application of Kirchhoff's voltage law around the emitter-biasing "loop" yields

$$V_{BB} - I_B R_B - v_{BE} = 0 \tag{13-10}$$

For a forward-biased silicon transistor, v_{BE} is approximately 0.7 V. Under this assumption,

$$I_B \cong \frac{V_{BB} - 0.7}{R_B} \tag{13-11}$$

Table 13-1 Transistor Nomenclature

	Emitter-base voltage	Collector-base voltage	Emitter, collector currents
Instantaneous total value	v_{EB}	v_{CB}	i_E, i_C
Instantaneous signal component	v_{eb}	v_{cb}	i_e, i_c
Quiescent or dc value	V_{EB}	V_{CB}	I_E, I_C
Effective (rms) value of signal	V_{eb}	V_{cb}	I_e, I_c
Supply voltage (magnitude)	V_{BB}	V_{CC}	

With a signal current i_i applied, the base current becomes

$$i_B = I_B + i_i \tag{13-12}$$

an ac component superimposed on a dc component.

By Kirchhoff's voltage law around the output "loop,"

$$v_{CE} = V_{CC} - i_C R_L \tag{13-13}$$

This is the equation of the load line shown in Fig. 13.17b. The quiescent point lies at the intersection of the load line and the characteristic curve for $i_B = I_B$. The intersections of the load line and the characteristic curves of the nonlinear transistor represent graphical solutions of Eq. 13.13 and define the instantaneous values of i_C and v_{CE} corresponding to values of the input signal current i_i. The signal output i_o is the sinusoidal component of i_C, the current in R_L. The current amplification, or current gain, is

$$A_I = \frac{i_o}{i_i} = \frac{I_{om}}{I_{im}} = \frac{I_o}{I_i} \tag{13-14}$$

depending on whether the current ratio is expressed in terms of instantaneous, maximum, or rms values. (See Example 5.)

Integrated Circuits

In 1958 J. S. Kilby developed an "integrated circuit," a single monolithic chip of semiconductor in which active and passive circuit elements were fabricated by successive diffusions and depositions. Shortly thereafter Robert Noyce fabricated a complete circuit including the interconnections on a single chip. The ability to control precisely the dimensions and doping concentration of diffused or deposited regions combined with advances in photolithographic techniques have made possible the mass production of sophisticated devices of very small size but exceedingly high reliability.

Example 5

The transistor whose characteristics are given below is used in the circuit of Fig. 13.17a. The Q point is to be at $I_B = 0.2$ mA, and $V_{CE} = 5$ V. Supply batteries available provide $V_{BB} = 3$ V, and $V_{CC} = 10$ V. Specify R_B and R_L and determine the current gain and output voltage for an input current $i_i = 0.1 \sin \omega t$ mA.

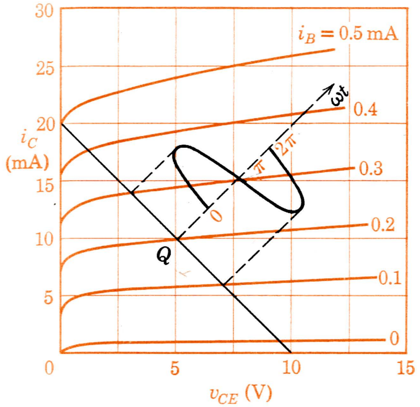


Figure 13.18 Amplifier performance.

By Eq. 13-11, the biasing resistance should be

$$R_B \cong \frac{V_{BB} - 0.7}{I_B} = \frac{3 - 0.7\text{ V}}{0.2\text{ mA}} = 11.5\text{ k}\Omega$$

For quiescent operation at $I_B = 0.2$ mA and $v_{CE} = V_{CE} = 5$ V, $I_C = 10$ mA. By Eq. 13-13,

$$R_L = \frac{V_{CC} - V_{CE}}{I_C} = \frac{10 - 5\text{ V}}{10\text{ mA}} = 500\ \Omega$$

The resulting load line is as drawn in Fig. 13.18.

For $\omega t = \pi/2$, the signal current is 0.1 mA, the base current is $I_B + i_i = 0.2 + 0.1 = 0.3$ mA, and the corresponding collector current is 14 mA. A signal current swing of 0.1 mA causes a collector current swing of 4 mA. The current gain is

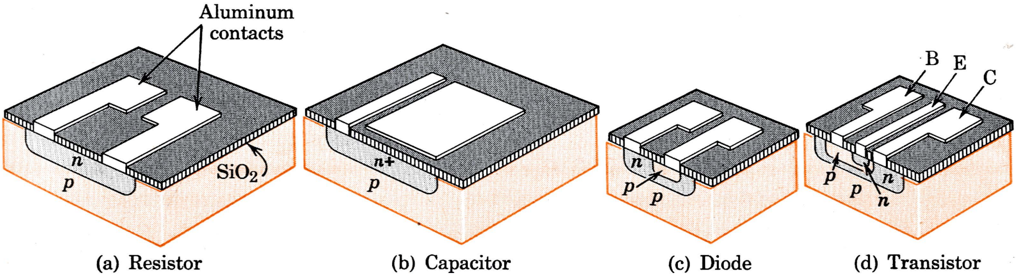
$$A_I = \frac{I_{om}}{I_{im}} = \frac{4\text{ mA}}{0.1\text{ mA}} = 40$$

The sinusoidal component of voltage across R_L is

$$V_{om} = I_{om} R_L = 0.004 \times 500 = 2\text{ V}$$

Since the input voltage variation associated with the input current is of the order of 0.1 V (see Fig. 13.15a), there also is an appreciable voltage gain.

With the exception of inductance, all the ordinary electronic circuit elements can be fabricated in semiconductor form. A resistor (Fig. 13.19a) is obtained in the form of a thin filament of specified conductivity, terminated in metallic contacts and isolated by a reverse-biased junction. Resistances from a few ohms to 20,000 ohms are possible. A low-loss capacitor is formed by depositing a thin



(a) Resistor (b) Capacitor (c) Diode (d) Transistor

Figure 13.19 Basic components of microcircuits.

insulating layer of silicon dioxide on a conducting region and then providing a metalized layer to form the second plate of the capacitor (Fig. 13.19b). Alternatively, a reverse-biased diode may be used to provide capacitance; capacitances up to 50 pF are possible. Typically, capacitors require the largest areas, resistors next, and active devices such as diodes, transistors, and FETs the smallest. Interconnections are provided by a layer of aluminum that is deposited, masked, and etched to leave the desired pattern.

Design Implications

The availability of sophisticated devices of high reliability in small packages at low cost has changed the character of electronic design. Instead of developing optimum circuits, buying components, and fabricating such units as amplifiers and digital registers, the design engineer selects mass-produced, high-performance functional units and incorporates them into his specific products. Originally conceived as a means for reducing the size of complex equipment, ICs now are equally noted for their low cost and high reliability.†

Since processing cost is proportional to area, circuits are designed to use more *active devices* (transistors and diodes) and fewer resistors and capacitors. (This is in sharp contrast to vacuum tube technology, where the active devices are the most expensive items by far.) The small circuit dimensions decrease the likelihood of pickup of unwanted noise; therefore signal levels can be lower. This in turn means lower voltages and lower power requirements. On the other hand, IC component values are hard to control precisely, and they may be voltage or temperature dependent. Also, unintentional capacitances may provide undesired coupling between circuit components.

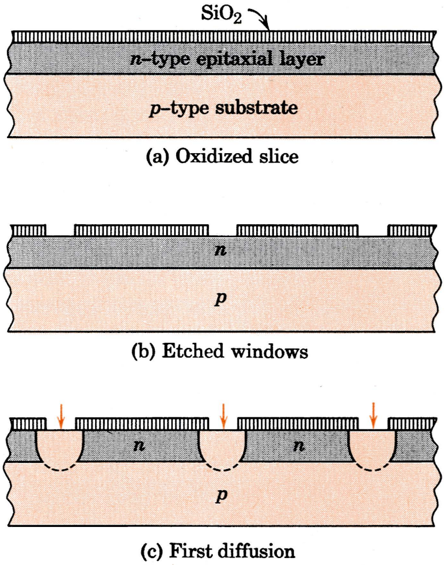
Commercial integrated circuits are usually available in three standard packages. The *multipin circular* type is the same size as a discrete transistor unit, but the chip it holds may contain hundreds of transistors and diodes. For use where size is important, the hermetically sealed *flat pack* occupies only 0.002 in.^{3} For convenience in use with printed-circuit boards, the *dual-in-line* package (DIP) provides easy plug-in operation. The cost of ICs ranges from less than a dollar to several dollars. In some cases, the cost of the package is greater than that of the chip.

Fabrication Processes

The fabrication of monolithic integrated circuits follows the planar technology used in transistor manufacture (Fig. 13.8b). A thin slice or *wafer* of p-type silicon 2 to 4 in. in diameter and 0.01 in. (10 mils) thick provides the *substrate.* The active and passive components are built within a thin n-type *epitaxial layer* on

---

† See References at the end of this chapter.


SiO_2
n-type epitaxial layer
p-type substrate
n
p
n
n
p

(a) Oxidized slice

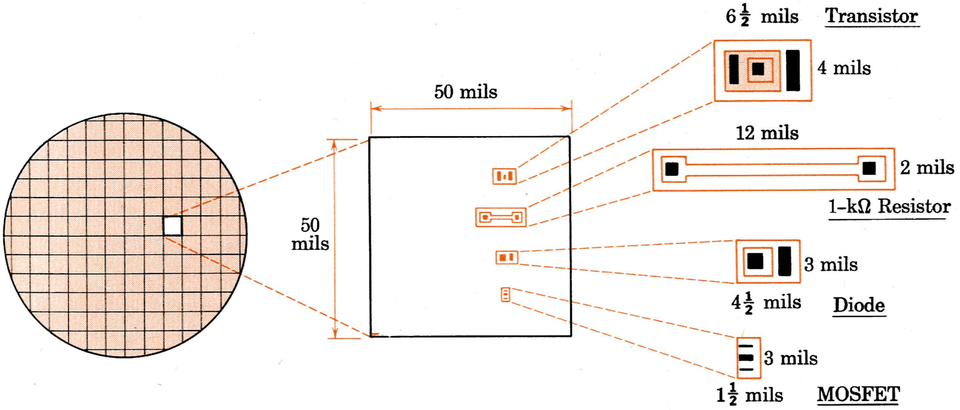
(b) Etched windows

(c) First diffusion

Figure 13.20 Insulation of IC elements by diode isolation.

top (Fig. 13.20a). First the surface is oxidized by heating it in the presence of oxygen. Then the cooled wafer is coated with a photosensitive material called *photoresist* and covered with a mask of the desired geometry. Upon exposure, the light-sensitive material hardens and the unexposed material is dissolved away. An etching solution, in which the photoresist is insoluble, is then used to remove the SiO_2, leaving a pattern of windows through which the first diffusion takes place.

One method of insulating components from each other is to provide *diode isolation* by surrounding each element with a reverse-biased *pn* junction. The first diffusion provides *p*-type regions that extend through to the substrate (Fig. 13.20c). The individual circuit elements are then built within these isolated islands by successive steps involving oxidation, masking, etching, and diffusion of *p*- and *n*-type impurities. In the final step, aluminum is evaporated onto the wafer and selectively removed to provide interconnections and relatively large *bonding pads* to which leads are welded.


50 mils
50 mils
$6\frac{1}{2}$ mils
Transistor
4 mils
12 mils
2 mils
1-kΩ Resistor
3 mils
$4\frac{1}{2}$ mils
Diode
3 mils
$1\frac{1}{2}$ mils
MOSFET

(a) Scribed silicon slice (b) Single IC chip (c) Typical dimensions

Figure 13.21 Typical integrated circuit details. (After R. G. Hibberd.)

Integrated circuits are cheap because thousands of complex units can be fabricated simultaneously. After the circuit is designed functionally and tested experimentally, it is designed dimensionally and masks are drawn about 500 times full scale. Then each mask is reduced photographically to actual size and precisely replicated to provide a master mask containing hundreds of identical patterns. Each processed slice (Fig. 13.21) is *diced* to yield several hundred individual chips which are tested and encapsulated. If each of 500 chips contains 50 components, and if 20 wafers are processed in a batch, a total of $500 \times 50 \times 20 = 500{,}000$ components are fabricated at once. Even if the yield of perfect chips is only 10%, the cost per component is very small.

Component Formation

Bipolar transistors in IC form are obtained by successive diffusions of boron (p-type) and phosphorus (n-type). A high concentration of phosphorus is necessary to overcome the previous boron diffusion and to create a high conductivity n^+-type emitter (Fig. 13.22). Another n^+ area is created to provide a low-resistance contact with the collector. (To minimize the series resistance of the thin collector region, another n^+ layer is formed in the substrate before the epitaxial n layer is grown.)

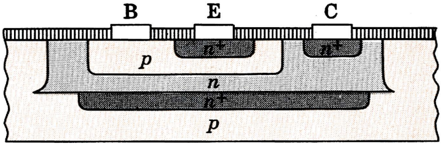


Figure 13.22 Structure of IC transistor.

The MOS transistor in IC form is identical with its discrete counterpart. MOS transistors can be smaller—and therefore cheaper—than bipolar transistors because they are self-isolating. The source, drain, and channel incorporate insulating pn junctions, and the gate is isolated by oxide.

Complementary devices, one n-channel and one p-channel, can be combined in a single unit. These CMOS units require only a single power supply, use very little power, and offer special advantages in digital circuits.

Diodes may be formed in various ways. For example, the emitter-base junction of a transistor (with collector tied to the base) provides a low-voltage diode with fast response. If the emitter is omitted, the remainder of a transistor provides a collector-base diode with higher reverse-voltage rating but slower response.

Resistors are usually formed during transistor base diffusion and the conductivity is dependent upon the transistor requirements. The desired value of resistance is obtained by specifying the length l and width w of the conducting path. Since $R = \rho l/A = \rho l/tw$, for a given resistivity ρ and thickness t,

$$R = \frac{\rho}{t} \cdot \frac{l}{w} = R_S \frac{l}{w} \tag{13-15}$$

where R_S is the *sheet resistance*, typically 100 to 200 Ω. (See Example 6.) Low values of resistance can be obtained by using the higher conductivity n^+-type emitter diffusion. Also, the channel between source and drain of an MOS transistor can be used as a resistor whose value is governed by gate voltage.

Example 6

Design a 4000-Ω resistor using stripes 0.5 mil wide with 0.5-mil spacing.

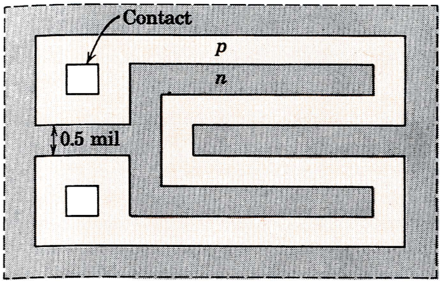


Figure 13.23 Layout of IC resistor.

Assuming $R_S = 100\ \Omega$, by Eq. 13-15,

$$l/w = R/R_S = 4000/100 = 40.$$

Therefore the conducting path should be 40 squares long.

For optimum space utilization, the path is folded back on itself. The stripe shown in Fig. 13.23 is approximately 40 squares long (end effects are neglected) and occupies an area of about 32 mils^2. (This is greater than the area required for a standard transistor.)

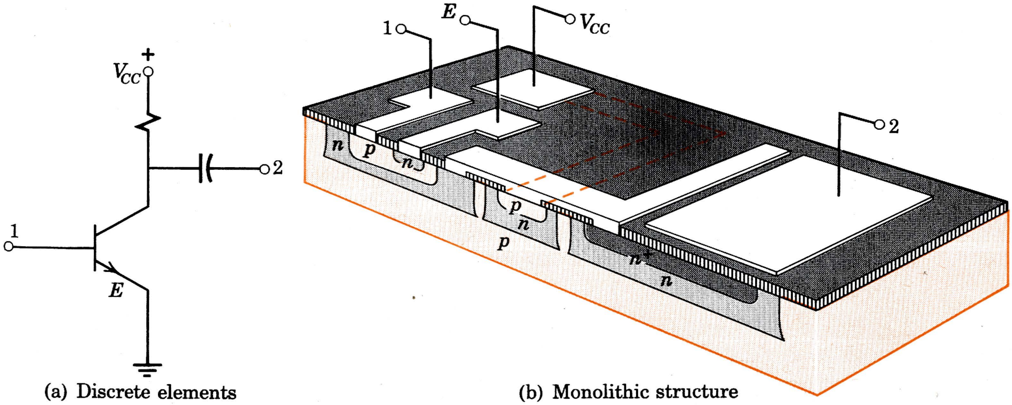
Capacitors may be fabricated by the MOS technique (Fig. 13.19b), or the inherent property of a reverse-biased *pn* junction can be used to provide a capacitor whose value is determined by bias voltage. Junction capacitors can be formed at the same time as the collector junctions of transistors, but the capacitance per unit area is quite low (~0.2 pF/mil^2). For MOS capacitors of area *A*, the capacitance is given by

$$C = \epsilon \frac{A}{t} \tag{13-16}$$

where ϵ = permittivity of silicon dioxide $\cong 3.5 \times 10^{-11}$ F/m. For a minimum practical thickness of $t = 0.002$ mil, the capacitance is about 0.4 pF/mil^2.

Applications

The design of integrated circuits consists in planning an optimum geometrical layout of standard components and connections (Fig. 13.24). Since the fabrication



(a) Discrete elements (b) Monolithic structure

Figure 13.24 An elementary amplifier in integrated circuit form (simplified).

of *microelectronic* devices requires sophisticated techniques and precision equipment, large volume production is necessary. Large volume means standardization on a relatively few products.

Integrated circuits are used in great quantities in digital computers because of their small size, low power consumption, and high reliability. The complexity of digital ICs ranges from simple logic gates and memory units (Chapter 14) to large arrays capable of complete data processing. The small size of MOS elements has led to *large scale integration* (LSI) in which thousands of elements are created on a single chip. Typical is the two-in-one chip developed for pocket calculators. This p-channel MOS device uses the equivalent of 3000 transistors and 400 logic gates to provide the four arithmetic functions and a memory unit.

In space vehicles and in hearing aids, the light weight of ICs is important. Microelectronics also finds application in device arrays where it is advantageous to have many matched elements on a single chip. For example, 20 of the amplifiers in Fig. 13.24 could be put on a single chip and, since they were fabricated simultaneously, they could be expected to "track" well with changes in temperature. More complex circuits, with high performance made possible by the liberal use of active devices, are used in the versatile *operational amplifiers* that are widely employed in place of individually designed units (Chapter 19).

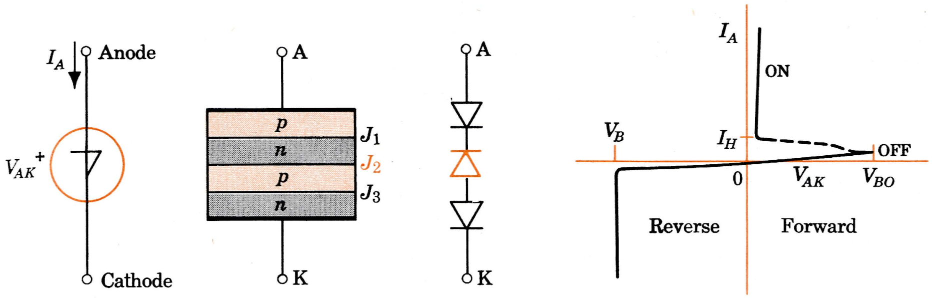
Thyristors

The diode is an automatic, unsophisticated switch; it is ON whenever the applied voltage provides forward bias. The transistor also can be used as a switch; collector current flows whenever there is adequate base current to turn the switch ON. Transistors are widely used as switches in digital circuits (Chapter 14) and in other cases where moderate amounts of power are to be controlled. A major disadvantage, however, is that transistors require high and continuous base current in the ON state. In contrast, the multilayered semiconductor devices called *thyristors* have the ability to control large amounts of power with only a minimum of control energy. As a result they are commonly used in sophisticated applications to rectification (ac to dc), inversion (dc to ac), relaying, timing, ignition, and speed control at power levels ranging from a few milliwatts to hundreds of kilowatts.

Operation of Four-Layer Devices

The silicon *pnpn* or *four-layer diode* consists of three *pn* junctions in series as shown in Fig. 13.25. The symbol contains a 4 with the apex pointing in the direction of positive current flow. In the three-diode representation, for negative anode-to-cathode voltage V_{AK}, center junction J_2 is forward biased, but the outside junctions are reverse biased. As a result, this is a "reverse blocking" diode and the characteristic for negative V_{AK}, the OFF state, resembles that of a reverse-biased diode. No reverse current flows unless the avalanche breakdown voltage is exceeded.

For positive V_{AK}, the current at low voltages is limited by the reverse bias behavior of junction J_2. As the applied voltage increases, the current increases


I_A
Anode
V_{AK}
+
Cathode
A
p
n
p
n
J_1
J_2
J_3
K
A
K
I_A
ON
V_B
I_H
OFF
0
V_{AK}
V_{BO}
Reverse
Forward

(a) Symbol (b) Cross section (c) Diode representation (d) Characteristic

Figure 13.25 The four-layer diode.

slowly until the *breakover voltage* V_{BO} is reached. At this point the current rises abruptly and the voltage drops sharply; the diode has switched to the ON state.

The switching operation of the four-layer diode can be understood by viewing the device as two interconnected transistors. In Fig. 13.26, back-to-back *pnp* and *npn* transistors are shown physically displaced but electrically connected. By Eq. 13-4, the collector currents will be

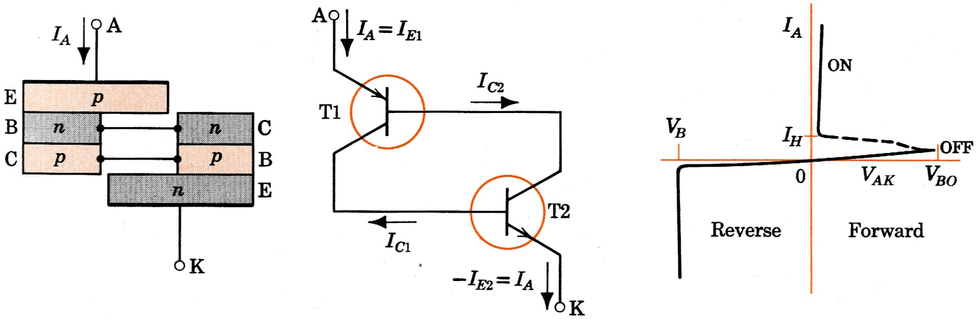
$$\begin{aligned} I_{C1} &= -\alpha_1 I_{E1} + I_{CBO1} = -\alpha_1 I_A + I_{CBO1} \\ I_{C2} &= -\alpha_2 I_{E2} + I_{CBO2} = +\alpha_2 I_A + I_{CBO2} \end{aligned} \qquad (13\text{-}17)$$

Summing the currents into transistor T1,

$$I_A + I_{C1} - I_{C2} = 0 = I_A - \alpha_1 I_A + I_{CBO1} - \alpha_2 I_A - I_{CBO2}$$

Solving, and recognizing that the cutoff currents of *pnp* and *npn* transistors are of opposite sign,

$$I_A = \frac{I_{CBO1} - I_{CBO2}}{1 - (\alpha_1 + \alpha_2)} = \frac{I_{CO}}{1 - (\alpha_1 + \alpha_2)} \qquad (13\text{-}18)$$


I_A
A
E
B
C
p
n
p
n
p
n
C
B
E
K
A
$I_A = I_{E1}$
T1
I_{C2}
T2
I_{C1}
$-I_{E2} = I_A$
K
I_A
ON
V_B
I_H
OFF
0
V_{AK}
V_{BO}
Reverse
Forward

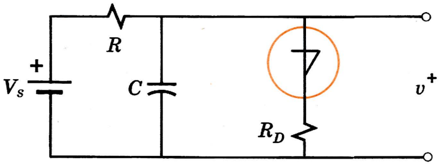
(a) Back-to-back transistors (b) Interconnected transistors (c) Characteristic

Figure 13.26 Analysis of four-layer diode operation.

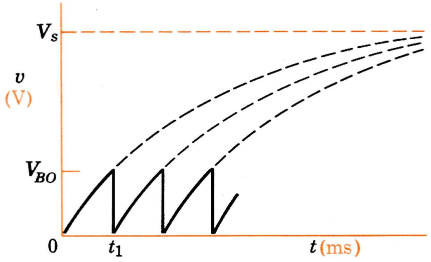
For silicon at very low emitter currents, α (see Fig. 13.10) is small because of recombination of holes and electrons in the transition region. As I_A increases with an increase in V_{AK}, α_1 and α_2 increase. As the quantity $(\alpha_1 + \alpha_2)$ approaches unity, the current tends to increase without limit and breakover occurs. After switching ON, the diode voltage is very small ($\sim$1 V) and the current is limited by the resistance of the external circuit. If the applied voltage is reduced, the switch remains ON until the current has reached a minimum *holding current* I_H. To switch the four-layer diode OFF, therefore, the current is reduced below I_H by reducing the applied voltage or increasing the series resistance.

Example 7

Design a sawtooth wave oscillator for operation at 500 Hz. Available is a four-layer diode rated at $V_{BO} = 100$ V, $I_A = 20$ A, and $I_H = 20$ mA.



(a) Circuit



(b) Output

Figure 13.27 Sawtooth oscillator.

The breakover characteristic of a *pnpn* diode makes it suitable for timing functions. In the circuit of Fig. 13.27a, the capacitor charges up to V_{BO} and discharges through the diode and resistor R_D.

For proper operation $V_s > V_{BO}$,

$$\therefore \text{ specify } V_s = 3\,V_{BO} = 3 \times 100 = 300 \text{ V.}$$

To ensure turnoff, $V_s/R < I_H$ or $R > V_s/I_H = 300/0.02 = 15\text{ k}\Omega$ $\therefore$ specify $R = 25\text{ k}\Omega$.

For $v/V_s = 1 - e^{-t_1/RC} = 100/300$ (Fig. 3.5),

$$t_1/RC \cong 0.4 = 1/RCf$$

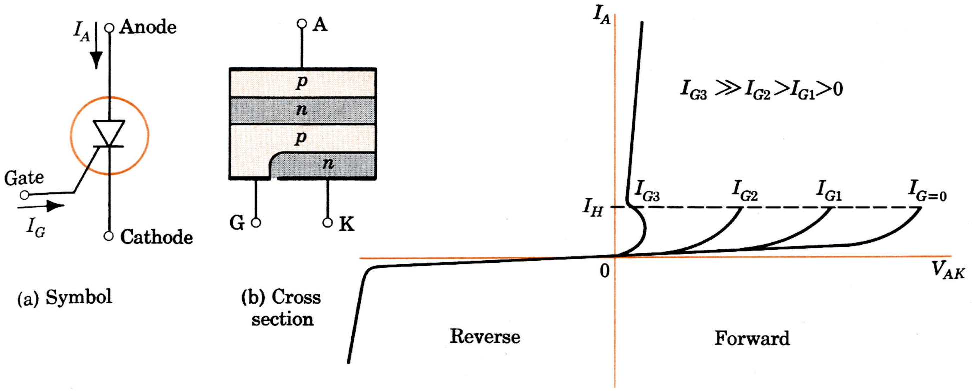
$$\therefore C = 1/0.4Rf = 1/0.4 \times 25 \times 10^3 \times 500 = 0.2\ \mu\text{F}$$

For $I_A = 20$ A, $R_D = V_{BO}/I_A = 100/20 = 5\ \Omega$.

The output (Fig. 13.27b) approximates a sawtooth.

Operation of Silicon Controlled Rectifiers

In a *pnpn triode* a *gate* connection to the inner *p* layer permits the introduction of a *gate current* that switches this thyristor to the ON state at voltages lower than V_{BO} for the corresponding diode. Such a *silicon controlled rectifier* (SCR) provides flexible switching and, since the gate current is required only momentarily, very efficient control of large amounts of power. In Fig. 13.28b, the gate connection is to the base of the *npn* transistor section. (In the *silicon controlled switch,* SCS, gate signals may be applied to the base of the *pnp* transistor as well.) Injection of base current increases α and reduces the necessary breakover


I_A
Anode
Gate
I_G
Cathode
A
p
n
p
n
G
K
I_A
$I_{G3} \gg I_{G2} > I_{G1} > 0$
I_H
I_{G3}
I_{G2}
I_{G1}
$I_{G=0}$
0
V_{AK}
Reverse
Forward

(a) Symbol

(b) Cross section

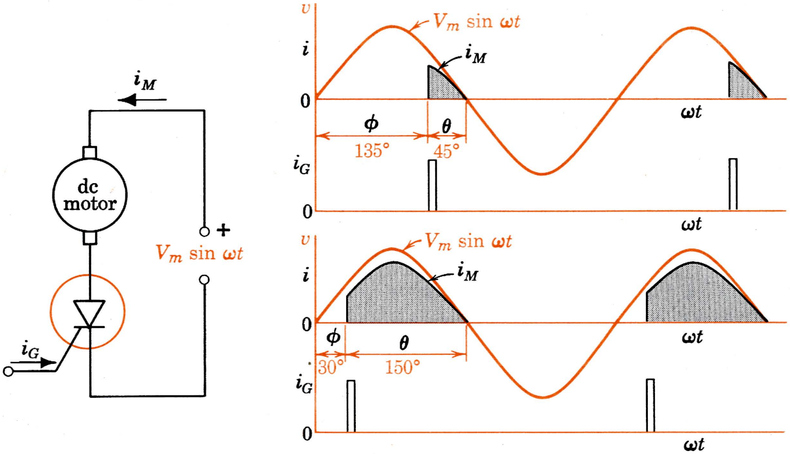
(c) I–V characteristic

Figure 13.28 The silicon controlled rectifier (SCR).

voltage. Therefore the firing voltage is a function of gate current I_G; for large values of I_G, the behavior approximates that of a simple *pn* diode.

In the typical application, firing is determined by a current pulse applied at the desired instant (Fig. 13.29). The *conduction angle* θ can be varied from zero to 180° by controlling the *firing angle* ϕ. Assuming that V_{AK} is negligibly small in the ON state and that motor current i_M is directly proportional to applied voltage v, the current waveform will be as shown. If the average or dc current is the important controlled variable, this can be determined from

$$I_{\text{dc}} = \frac{1}{2\pi}\int_{\phi}^{\pi} I_m \sin\omega t \, d(\omega t) = \frac{I_m}{2\pi}(1 + \cos\phi) = \frac{I_m}{2\pi}(1 - \cos\theta) \quad (13\text{-}19)$$


i_M
dc motor
+
$V_m \sin\omega t$
i_G
v
i
$V_m \sin\omega t$
i_M
0
ϕ
θ
135°
45°
ωt
i_G
0
ωt
v
i
$V_m \sin\omega t$
i_M
0
ϕ
θ
30°
150°
ωt
i_G
0
ωt

(a) Motor circuit

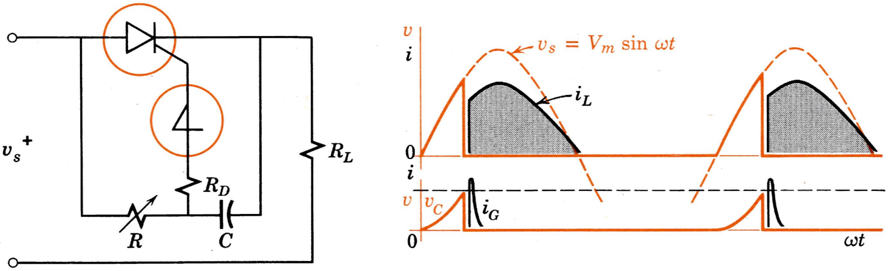
(b) Waveforms for $\theta = 45°$ and 150°

Figure 13.29 SCR speed control of a dc motor.

Note that once the SCR has turned ON, gate current is no longer required. The SCR is turned OFF when the supply voltage goes through zero.

SCR Ratings and Applications

In comparison to power transistors, thyristors are designed with thick base layers for high voltage capability and the necessary low α characteristic. Currents may be larger because the total junction areas are available for conduction. Maximum SCR ratings up to 2000 V and 600 A (average) are available. In common sizes, gate current pulses of a few milliamperes (50 mA) requiring gate voltages of a few volts (1–2 V) will control anode currents several thousand times as large (100 A). For a typical *turn-on* time of 1 μs, the average gate power is very, very small. In general, the *turn-off* time (during which the anode voltage must be kept below the holding voltage) is several times as long as the turn-on time. In comparison to other power rectifiers, thyristors are compact and extremely reliable; they have no inherent failure mechanism.†



(a) Circuit diagram (b) Voltage and current waveforms

Figure 13.30 SCR rectification with *pnpn* diode trigger.

The high power gain, efficiency, and reliability of thyristors have led to a host of applications. In the power control circuit of Fig. 13.30, an SCR provides rectification and a *pnpn* diode serves as a trigger timed by variable resistance R. With the SCR OFF, supply voltage v_s appears across RC, the capacitor begins to charge, and v_C increases at a rate determined by R. (Note the difference in response from that in the sawtooth oscillator of Fig. 13.27.) When v_C exceeds the breakover voltage of the *pnpn* diode, the capacitor discharges through current-limiting resistance R_D, providing a gate current pulse that fires the SCR. If full-wave rectification is desired, a bridge rectifier can be placed in front of the control circuit of Fig. 13.30a. For light dimming or ac motor speed control, two SCRs can be connected in inverse parallel to provide current on both half cycles. The *triac* (*tri*ode *ac* switch), a popular five-layer thyristor, is essentially an integrated double-ended SCR in a single unit.

---

† *Silicon Controlled Rectifier Manual,* 2nd ed., General Electric Co., Auburn, New York, 1961.

SUMMARY

- In a field-effect transistor, the gate voltage controls the channel conductivity.
  The mobile carrier density may be depleted or enhanced.
  In the constant-current region, transfer characteristics are parabolic.
  FETs are characterized by high input resistance.

- A junction transistor consists of two *pn* junctions in close proximity; normally, the emitter junction is forward biased, the collector reverse.
  In common-base operation,

$$i_C = -\alpha i_E + I_{CBO} \quad \text{where } \alpha \cong 1$$

  In common-emitter operation, a small base current controls the relatively larger collector current to achieve current amplification.

$$i_C = \beta i_B + I_{CEO} \quad \text{where } \beta = \frac{\alpha}{1-\alpha}$$

- Graphical analysis of nonlinear transistors is based on load-line construction.
  Variations in input quantities about a properly selected quiescent point result in corresponding output variations.

- Low cost ICs have changed the character of electronic design.
  Transistors, diodes, resistors, and capacitors are fabricated simultaneously by successive oxidation, masking, etching, and diffusion steps.
  For diffused resistors (up to 20 kΩ) $R = R_S l/w$.
  For MOS capacitors (up to 100 pF) $C = \epsilon A/t$.

- Thyristors control large amounts of power with minimum control energy.
  The *pnpn* diode turns ON when the breakover voltage is exceeded.
  The SCR triode thyristor can be turned ON by timed gate current pulses.
  Because they are compact, efficient, and reliable, SCRs are widely used in rectification, inversion, timing, relaying, and power control.

REFERENCES

1. Robert Hibberd, *Integrated Circuits*, McGraw-Hill Book Co., New York, 1969.
   An excellent introduction to IC design, technology, and applications; for engineers and technicians.
2. Lothar Stern, *Fundamentals of Integrated Circuits*, Hayden Book Co., New York, 1968.
   A more technical and more detailed treatment of IC technology and applications; for engineers.
3. Leonce Sevin, *Field-Effect Transistors*, McGraw-Hill Book Co., New York, 1965.
   The theory, physical behavior, and application of FETs; a fairly advanced treatment.

REVIEW QUESTIONS

1. Explain the function of the source, gate, and drain in the junction gate FET.
2. Sketch the cross section of a JFET at pinch off.
3. Explain the operation of an enhancement-type MOSFET.
4. Explain with sketches the operation of a *pnp* transistor.
5. Why must the base be narrow for transistor action?
6. Describe how amplification is achieved by a transistor.
7. Sketch, from memory, the characteristics of an ideal current amplifier; superimpose the characteristics of a common-emitter transistor.
8. Explain common-base transistor characteristics in terms of diode characteristics.
9. Explain the role of the load line in graphical analysis of amplifiers.
10. List two ways in which the availability of ICs has changed electronic design.
11. Explain how system reliability is increased by using integrated circuits.
12. How many diffusion steps are required in typical IC fabrication?
13. How does diode isolation provide element insulation?
14. Without reference to the text, outline the process of making an IC including resistors, capacitors, and transistors.
15. What are the advantages of thyristors over *pn* diodes? over transistors?
16. Explain the behavior of a *pnpn* diode as breakover is approached and reached.
17. Why doesn't removal of gate current turn OFF an SCR?

EXERCISES

1. An electronic two-port is defined by the following equations:

$$\begin{cases} v_1 = 0.7 + 1500 i_1 \text{ V} \\ i_2 = 30 i_1 + 0.001 \text{ A} \end{cases}$$

A 2000-Ω resistance is connected across the output terminals (subscript 2), and a voltage $v_1 = 0.8 + 0.1 \sin \omega t$ V is applied to the input terminals (see Fig. 10.20a).

   (a) Sketch the input and output characteristics for this device for $0 < v_1 < 3$ V and $0 < i_1 < 0.25$ mA, respectively.

   (b) Determine the signal (sinusoidal) components of the input and output voltages, and the voltage amplification.

2. Consider the FET of Fig. 13.2 as an approximation to the ideal transducer of Fig. 10.21b. An average gate voltage $V_{GS} = -2$ V is maintained and a load resistance $R_L = 5$ kΩ is inserted in the drain circuit of Fig. 13.1c. Estimate the output (drain) current variation for an input (gate-source) signal of 1 sin ωt V and the voltage amplification possibilities of this amplifier.

3. A JFET constructed like the one in Fig. 13.1b has the following typical properties: $I_{DSS} = 4$ mA and $V_{GS(\text{OFF})} = -5$ V.

   (a) Sketch a set of i_D vs v_{DS} curves.

   (b) Draw transfer characteristic i_D vs v_{GS}.

   (c) For $v_{GS} = -2$ V in Fig. 13.31, specify R_S for operation in the normal region.

   (d) For the value of R_S in part (c) and $R_D = 3$ kΩ, select a reasonable value of v_{DS} and specify V_{DD}.

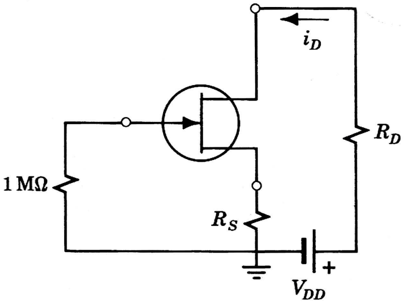


Figure 13.31

4. The MOSFET of Fig. 13.3 is to be used as an elementary amplifier.

   (a) Draw an appropriate circuit showing bias voltages V_{GG} and V_{DD} and load resistor R_L.

   (b) For $V_{DD} = 12$ V, $V_{GG} = 8$ V, and $R_L = .2$ kΩ, determine the quiescent drain current I_D.

(c) Estimate the input signal voltage for an output $v_o = 1 \sin \omega t$ V.

5. The MOSFET of Fig. 13.3 is to be used as a passive load resistor by connecting the gate terminal to the drain.
   (a) Draw a graph of i_D versus v_{DS} for this connection.
   (b) Estimate the resistance offered to a dc current of 2 mA.
   (c) Use piecewise linearization to represent this device by a combination of voltage source and resistance.
6. For a *pnp* transistor,
   (a) Sketch the charge distribution when unbiased.
   (b) Sketch the potential distribution when unbiased.
   (c) Sketch the potential distribution when properly biased.
7. For high-gain amplification, high values of α and β are desirable. Considering the effect of doping on factor γ (Fig. 13.10), explain why it is desirable to dope the base relatively lightly in comparison to the emitter and collector regions.
8. Assume that your roommate has a good knowledge of physics, including semiconductors, and mathematics, but has never studied electronics. Using labeled sketches and graphs, explain to him or her the "normal" operation of a *pnp* junction transistor.
9. A transistor is connected as in Fig. 13.32.

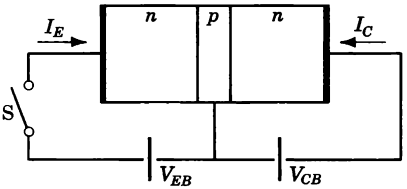


Figure 13.32

   (a) Label the polarities of v_{EB} and v_{CB} to provide for *normal* operation.
   (b) For switch S *open*, sketch a graph of I_C versus $\pm V_{CB}$.
   (c) For switch S *closed* and $I_E = -2$ mA, repeat part (b).
10. A silicon *pnp* transistor is connected in the circuit shown in Fig. 13.33.

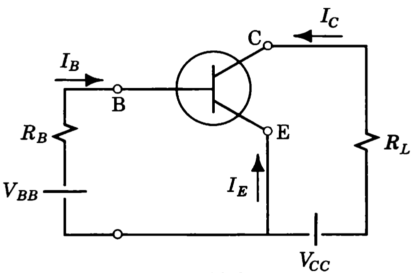


Figure 13.33

   (a) Complete the drawing by putting the arrowhead on the emitter symbol and indicating the proper battery polarities for a forward-biased emitter-base junction and a reverse-biased collector-base junction.
   (b) Assuming $|V_{BE}| = 0.7$ V, calculate I_B for $V_{BB} = 4$ V and $R_B = 33$ kΩ.
   (c) If $V_{CC} = 15$ V, $R_L = 1.5$ kΩ, and $\beta = 60$, calculate I_C, I_E, and V_{CE}.
11. A 2N007 (a rather poor *pnp* transistor) has the following parameters: $\alpha = 0.97$ and $I_{CBO} = 10\ \mu$A. For $0 < i_C < 10$ mA,
   (a) Sketch the common-base collector characteristics.
   (b) Sketch the common-emitter collector characteristics.
12. Consider the transistor of Fig. 13.15b as an approximation to the idealized current amplifier of Fig. 10.20. If an average collector current of 10 mA is maintained, estimate the current amplification possibilities of this device.
13. A certain IC device would cost \$20,000 for the complete design and \$1 each to manufacture. Predict:
   (a) The unit cost for a custom run of 100.
   (b) The unit cost for a standard run of 100,000.
   (c) The number of units at which the design cost is 10% of the manufacturing cost.
14. An electronic computer contains 50,000 active devices.
   (a) If, on the average, the computer is to operate a week without component failure, what must be the life expectancy of each component?
   (b) If ICs containing 100 active components have the same reliability as the discrete components of part (a), what must be the life expectancy of the ICs for the same computer reliability?
15. Sketch the masks required to fabricate an IC transistor.

16. Assuming $R_S = 200\ \Omega$, lay out and determine the dimensions of a 10-kΩ diffused *p*-type resistor for an IC. Use 0.5-mil stripes with 0.5-mil spacing in an approximately square area.
17. Lay out and determine the dimensions of a 10-pF IC capacitor, assuming an oxide thickness of 0.003 mil.
18. In Fig. 13.27a, $V_s = 200$ V, $R = 10$ kΩ, and $C = 0.5\ \mu$F. If V_{BO} is 50 V, predict the oscillator frequency.
19. Sketch waveforms of $v_S = V_S \sin \omega t$, $v_G = V_G \sin (\omega t - 45°)$, i_L, and v_{AK} for an SCR in series with v_S and R_L.
20. Redraw Fig. 13.26b to represent an SCR by interconnected transistors. Express I_{C1} and I_{C2} in terms of β_1 and β_2, and derive an expression for I_A. What is the condition for turn-on? Is this consistent with the conclusion reached for the *pnpn* diode?
21. In Fig. 13.30, $V_m = 200$ V and $R_L = 10\ \Omega$.
    (a) For a conduction angle of 120°, determine average load current and power.
    (b) Repeat for a conduction angle of 75°.
    (c) If the ON voltage across the SCR is 1 V, calculate the power lost in the diode in part (a).
22. In Fig. 12.9a, $V_m = 300$ V and $R_L = 20\ \Omega$. It is desired to control the power delivered to the load without changing v or R_L.
    (a) Redraw the circuit showing how an SCR could be used for this purpose.
    (b) Sketch voltage and current waveforms for an SCR conduction angle of 45°.
    (c) Predict the power delivered to R_L for conduction angles of 45° and 135°.

PROBLEMS

1. The MOSFET of Fig. 13.3 is to be used as a switch in a digital computer. The drain is connected through a resistance $R_D = 6$ kΩ to a voltage $V_{DD} = 12$ V, and source S is grounded. V_{GS} is the input and V_{DS} is the output.
    (a) Draw the circuit diagram.
    (b) Estimate the output voltage for inputs of −2 and −10 V.
    (c) Estimate the "open" and "closed" resistances for this switch.
2. In a conventionally arranged amplifier, the transistor terminals are labeled *X*, *Y*, *Z*. With no signal applied, it is observed that terminal *X* is 20 V positive with respect to terminal *Y*, and 0.5 V positive with respect to terminal Z. Explain your reasoning:
    (a) Which terminal is the emitter?
    (b) Which terminal is the collector?
    (c) Which terminal is the base?
    (d) Whether *X* is *p*- or *n*-type material?
3. The variation in α with v_{CB} in a transistor (Fig. 13.12c) can be explained in terms of the effect of v_{CB} on the *effective width* of the base.
    (a) As the reverse bias increases, what is the effect on the width of the transition (or depletion) region adjacent to the base-collector junction?
    (b) How does a change in transition width change the effective base width?
    (c) How does base width affect α?
    (d) Why does an increase in v_{CB} produce the effect shown in Fig. 13.12b?
4. A transistor amplifier using a 2N699B is to supply a 10-V (rms) output signal across a 10-kΩ load resistance. "Design" an elementary amplifier by drawing an appropriate circuit and specifying the components. Estimate the input signal current required.
5. A subcircuit commonly used in integrated circuits is shown in Fig. 13.34. (Assume $\beta = 30$ for the transistor.)
    (a) Sketch the *i-v* characteristic of a silicon diode and estimate the voltage at point *x* with respect to point *y* if the diode is conducting.
    (b) If the base-emitter junction of the transistor is forward-biased, estimate the voltage of point *y* and the voltage at point *x* with respect to ground.
    (c) Estimate the base and collector currents of the transistor and the output voltage V_o under these conditions.

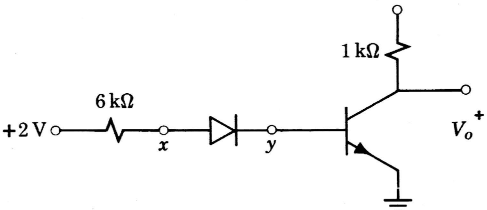


Figure 13.34

Chapter 14

Digital Devices

Switching Logic
Electronic Switches
Memory Devices

The audible portion of a program transmitted by a modulated television carrier is in the form of *continuous* or *analog* signals; the signal amplitude or frequency can take any value within a wide range of values. In contrast, the information that triggers the scanning of the picture tube is in the form of *discrete* or *digital* signals; the trigger signal is present or absent. A major virtue of electronic circuits is the ease and speed with which digital signals can be processed, and the use of such signals in control, computation, and instrumentation is the most rapidly developing aspect of electrical engineering.

The processing of information in digital form requires special circuits, and the efficient design of digital circuits requires a special numbering system and even a special algebra. The circuits must provide for storing instructions and data, receiving new data, making decisions, and communicating the results. For example, an automatic airline reservation system must receive and store information from the airline regarding the number of seats available on each flight, respond to inquiries from travel agents, subtract the number of seats requested from the number available or add the number of seats canceled, handle 50 or so requests per minute, and keep no one waiting more than a minute.

Information processing is an important component of all branches of engineering and science. Aeronautical and chemical engineers may be designing automatic control systems. Civil and industrial engineers may be concerned with data on traffic flow or product flow. Chemists and medical doctors may be interested in automated laboratory analysis. Biologists and physicists may need remote control of precisely timed events. Everyone engaged in experimental work can benefit from the new data processing techniques.

In this chapter, only a brief look at this important field can be provided. The design of practical circuits for optimum performance is left to advanced courses for students with special interests. The emphasis here is on understanding the concepts of switching logic and seeing how semiconductor devices can be used in decision-making and memory functions.

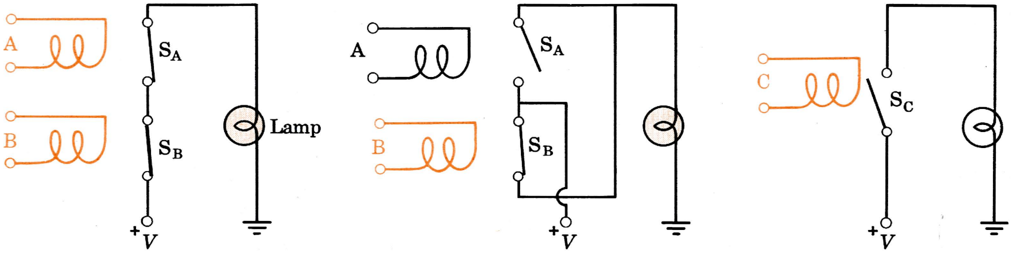
Switching Logic

Digital computers, automatic process controls, and instrumentation systems have the ability to take action in response to input stimuli and in accordance with instructions. In performing such functions, an information processing system follows a certain *logic;* the elementary logic operations are described as **AND**, **OR**, and **NOT**. Tiny electronic circuits consuming very little power can perform such operations dependably, rapidly, and efficiently.

Gates

A *gate* is a device that controls the flow of information, usually in the form of pulses. First we consider gates employing magnetically operated switches called *relays*. If the switches are normally open, they close when input signals in the form of currents are applied to the relay coils.

In Fig. 14.1a, the lamp is turned on if switch A *and* switch B are closed; it is therefore called an **AND** circuit or **AND** gate. In Fig. 14.1b, the lamp is turned on if switch A *or* switch B is closed *or* if both are closed; it is, therefore, called an **OR** circuit. In Fig. 14.1c, the switch is normally closed; for an input at C there is **NOT** an output or the input has been *inverted.*



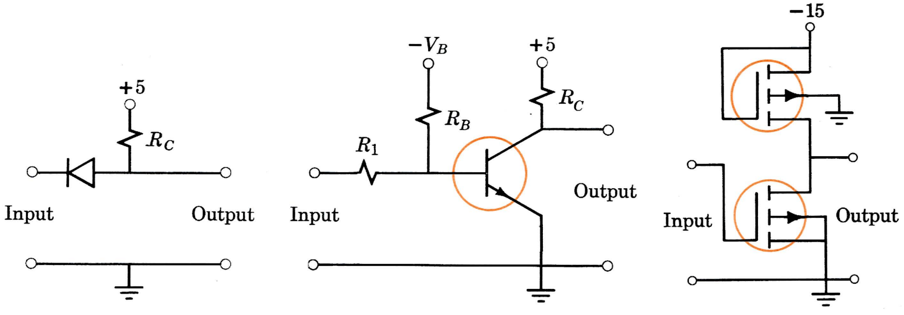
(a) **AND** gate (b) **OR** gate (c) **NOT** gate

Figure 14.1 Gate circuits using relays.

In general, there may be several inputs and the output may be fed to several other logic elements. In a typical digital computer there are millions of such logic elements. Although the first practical digital computer (the Harvard Mark I, 1944) employed relays, the success of the modern computer is due to our taking advantage of the switching capability of semiconductor devices.

Electronic Switching

In the gates of Fig. 14.1, "data" are represented by ON and OFF switch positions or by the corresponding presence or absence of voltages. How can diodes and transistors be used as switches? If the input to the circuit in Fig. 14.2a is zero (i.e., if the input terminals are shorted), the diode is forward biased, current flows



(a) Diode switch (b) Transistor switch (c) MOS IC switch

Figure 14.2 Elementary semiconductor switching circuits.

easily, and the output is a few tenths of a volt or approximately zero. If the input is 5 V or more, no diode current flows, and the output is +5 V. The OFF–ON positions of this diode switch correspond to voltage levels near zero and +5 V.

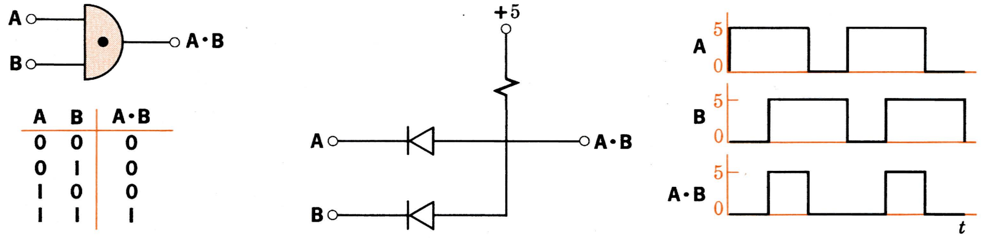
In the elementary transistor switch (Fig. 14.2b) with a zero input, the emitter junction is reverse biased by V_B, the collector current is very small, and the output is approximately +5 V. A positive input forward biases the emitter junction, a large collector current flows, and the output is approximately zero. Again the output voltage levels are near zero and +5, but the signal has been inverted in the transistor switch.

The MOS transistor can operate as a binary switch controlled by changes in gate voltage. Since an MOS transistor can also function as a resistor, the arrangement of Fig. 14.2c is convenient in integrated circuits. With the gate of the upper transistor connected to the drain terminal, enhancement is high and the resistance of the "load" is determined by the channel dimensions. With no input signal, the lower transistor is OFF and the output is approximately −15 V. A negative input signal turns the "driver" transistor ON and the output rises to nearly zero.

Basic Logic Operations

To facilitate our work with logic circuits, we use some of the conventional symbols and nomenclature developed by circuit designers and computer architects. Each basic logic operation is indicated by a *symbol* and its function is defined by a *truth table* that shows all possible input combinations and the corresponding outputs. Logic operations and variables are in boldface type. Electronic circuits are drawn as single-line diagrams with the ground terminal omitted; all voltages are with respect to ground so that "no input" means a shorted terminal. The two distinct voltage levels provided electronically represent the binary numbers **1** and **0** corresponding to the TRUE and FALSE of logic.

AND *Gate.* The symbol for an AND gate is shown in Fig. 14.3, where A · B is read "A AND B." As indicated in the truth table, an output appears only when there are inputs at A AND B. If the inputs are in the form of positive voltage pulses (with respect to ground), inputs at A and B reverse bias both diodes, no current flows through the resistance, and there is a positive output (1). In general, there may be several input terminals. If any one of the inputs is zero (0), current flows through that forward-biased diode and the output is nearly zero (0). For two inputs varying with time, a typical response is shown in Fig. 14.3c.

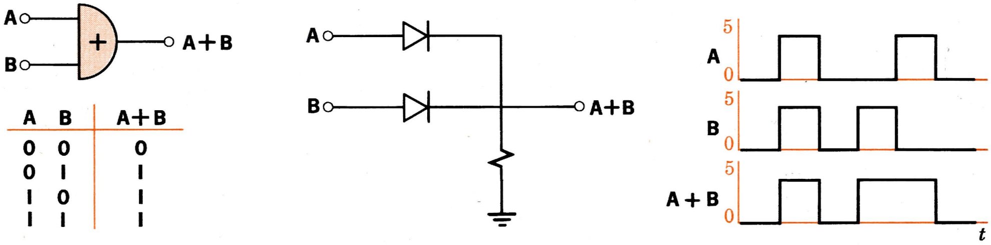


A	B	A·B
0	0	0
0	1	0
1	0	0
1	1	1

(a) Symbol and truth table (b) Diode circuit (c) Typical response

Figure 14.3 A two-input AND gate.

OR *Gate.* The symbol for an OR gate is shown in Fig. 14.4 where A + B is read "A OR B." As indicated in the truth table, the output is 1 if input A OR input B is 1. For no input (zero voltage) in Fig. 14.4b, no current flows, and the output is zero (0). An input of +5 V (1) at either terminal A or B or both (or at *any* terminal in the general case) forward biases the corresponding diode, current flows through the resistance, and the output voltage rises to nearly 5 V (1). For two inputs varying with time, the response is as shown.

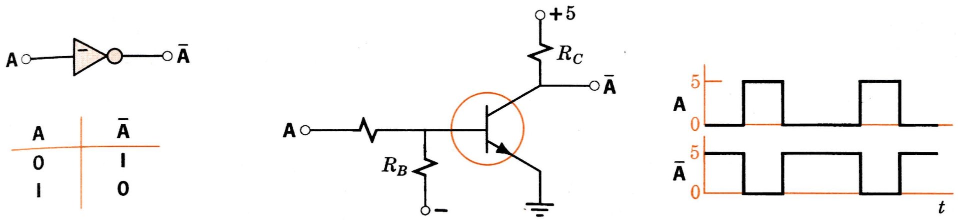


A	B	A+B
0	0	0
0	1	1
1	0	1
1	1	1

(a) Symbol and truth table (b) Diode circuit (c) Typical response

Figure 14.4 A two-input OR gate.

NOT *Gate.* The inversion inherent in a transistor circuit corresponds to a logic **NOT** represented by the symbol in Fig. 14.5 where $\bar{A}$ is read "**NOT A.**" As



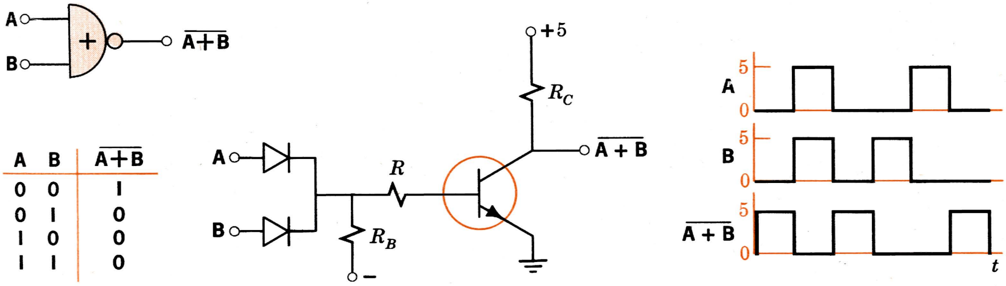
A	Ā
0	I
I	0

(a) Symbol and truth table (b) Transistor circuit (c) Typical response

Figure 14.5 A transistor **NOT** gate.

indicated in the truth table, the **NOT** element is an *inverter;* the output is the *complement* of the single input. With no input (**0**), the transistor switch is held open by the negative base voltage and the output is +5 V (**I**). A positive input voltage (**I**) forward biases the emitter junction, collector current flows, and the output voltage drops to a few tenths of a volt (**0**). For a changing input, the output response is as shown.

NOR *Gate.* In the diode **OR** gate of Fig. 14.4b, a positive input at **A OR B** produces a voltage across R and a positive output voltage. But this output is less than the input (by the diode voltage drop), and after a few cascaded operations the signal would decrease below a dependable level. A transistor can be used to restore the level (Fig. 14.6b); however, the inherent inversion results in



A	B	$\overline{A+B}$
0	0	I
0	I	0
I	0	0
I	I	0

(a) Symbol and truth table (b) Electronic circuit (c) Typical response

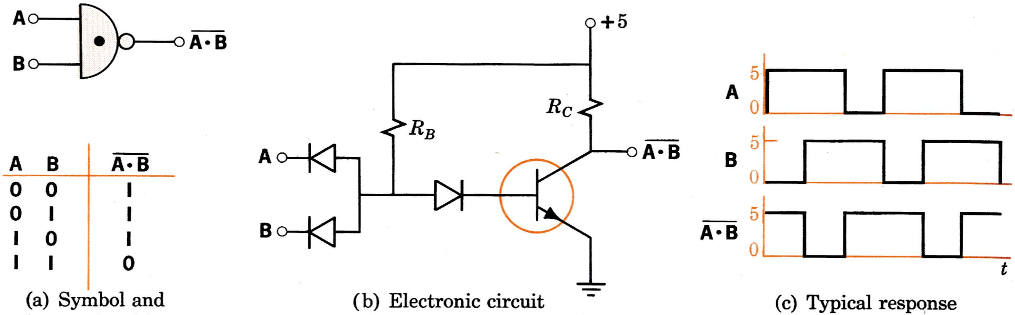
Figure 14.6 A diode-transistor **NOR** gate.

a **NOT OR** or **NOR** operation. The small circle on the **NOR** element symbol and the bar in the $\overline{A + B}$ output indicate the inversion process.

In the **NOR** circuit with no input, the transistor switch is held OFF by the negative base voltage and the output is +5 V (**1**). A positive voltage at terminal A or B raises the base potential, forward biases the emitter junction, turns the transistor switch ON, and drops the output to nearly zero (**0**). In addition to restoring the signal level, the transistor provides a relatively low output impedance so that this **NOR** element can supply inputs to many other gates. Another advantage is that all the basic operations can be achieved using only **NOR** gates.

NAND *Gate.* Diodes and a transistor can be combined to perform an inverted **AND** function. Such a **NAND** gate has all the advantages of a **NOR** gate and is very easy to fabricate, particularly in integrated circuit form. In a complex logic system it is convenient to use just one type of gate, even when simpler types would be satisfactory, so that gate characteristics are the same throughout the system.

The **NAND** gate function is defined by the truth table in Fig. 14.7. The small circle on the **NAND** element symbol and the bar on the $\overline{A \cdot B}$ output indicate the inversion process. With positive inputs at A and B (**1 AND 1**), the diodes are reverse biased, and no diode current flows; the positive base current causes heavy collector current and the output is approximately zero (**0**). If either A or B has a zero (**0**) input, at least one diode conducts to ground, the emitter junction voltage drops below the critical value, no base current and hence no collector current flows, and the output is +5 V (**1**). For more positive cutoff, a second diode may be placed in series with the base of the transistor.



A	B	$\overline{A \cdot B}$
0	0	1
0	1	1
1	0	1
1	1	0

(a) Symbol and truth table (b) Electronic circuit (c) Typical response

Figure 14.7 A diode-transistor **NAND** gate.

In Example 1, we see that three **NAND** gates could be used to replace an **OR** gate. The combination of **NAND** gates is equivalent to an **OR** gate in that it performs the same logic operation. In digital nomenclature, the function **f** is defined by

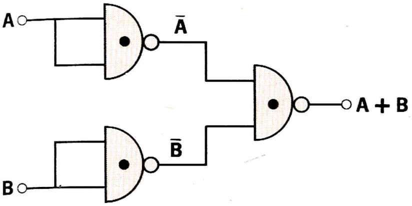
$$f = A + B = \overline{\overline{A} \cdot \overline{B}}$$

Example 1

Use **NAND** gates to form a two-input **OR** gate.

A	B	f	$\bar{A}$	$\bar{B}$	$\overline{\bar{A}\cdot\bar{B}}$
0	0	0	1	1	0
0	1	1	1	0	1
1	0	1	0	1	1
1	1	1	0	0	1

(a) Desired function



(b) **NAND** realization

Figure 14.8 Using **NAND** gates to form an **OR** gate.

The desired function is defined by the truth table of Fig. 14.8a. Comparing this with Fig. 14.7a, we see that if each input were inverted (replaced by its complement), the **NAND** gate would produce the desired result as indicated in the truth table.

To provide simple inversion, we tie both terminals of a **NAND** gate together (the first and last rows of Fig. 14.7a). The desired logic circuit is shown in Fig. 14.8b.

We arrived at this relation by considering the desired and available truth tables. A "digital algebra" for the direct manipulation of such expressions is presented in Chapter 15. First, however, let us see how practical semiconductor logic elements function.

Electronic Switches

Binary digital elements must respond to two-valued input signals and produce two-valued output signals. In practice the "values" are actually ranges of values separated by a *forbidden region*. The two discrete *states* may be provided electrically by switches, diodes, or transistors. They may be provided also by relays, or they may be provided optoelectronically by means of phototubes (Fig. 10.18) or semiconductor photo diodes. Magnetic cores in which the two states are represented by opposite directions of magnetization are used extensively in computers. In one interesting development, purely fluid switches are being used in control applications. In applications requiring high speed and flexibility, however, electronic switches predominate.

Classification of Electronic Logic

Electronic logic circuits are classified in terms of the components employed. Among the more common are diode logic (DL), resistor-transistor logic (RTL), diode-transistor logic (DTL), transistor-transistor logic (TTL), and emitter-coupled logic (ECL). In specifying a logic system, the designer selects the type of logic whose characteristics match the requirements.

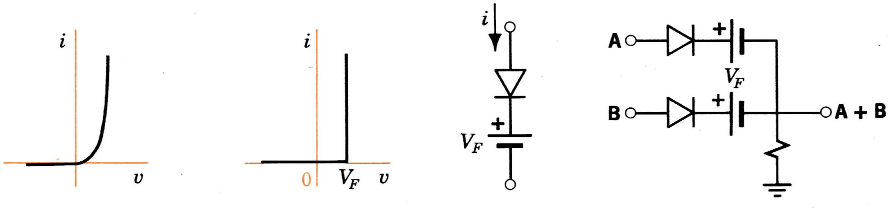
Logic types vary in *signal degradation, fan-in, fan-out,* and *speed.* A major disadvantage of diode logic (Figs. 14.3 and 14.4) is that the forward voltage drop is appreciable (and varies among diodes of the same type), and the output signal is "degraded" in that the forbidden region is narrowed. The use of transistors minimizes degradation. The number of inputs that can be accepted is called the fan-in and is low (3 or 4) in DL and high (8 or 10) in DTL. The number of inputs that can be supplied by a logic element is called the fan-out. Fan-out depends on the output current capability (and the input current requirement) and varies from 4 in DL to 10 or more in TTL.

The speed of a logic operation depends on the time required to change the voltage levels, which is determined by the effective time constant of the element. In high-speed diodes, the charge storage is so low that response is limited primarily by wiring and load capacitances. In transistors in the ON state, base current is high and the charge stored in the base region is high; before the collector bias can reverse, this charge must be removed. Typically, 5 to 10 nanoseconds are required to process a signal. (In ECL the charge stored is minimized and ECL gates can operate at rates up to 200 MHz.)

In practical design, the important factors are cost, speed, immunity to noise, power consumption, and reliability. In integrated circuit manufacture, a gate containing ten highly uniform active components may cost no more than a discrete transistor and is no less reliable. As a result, sophisticated logic elements of greatly improved characteristics are now available at reasonable cost.

Diode Switches

The diode is an automatic "gate," CLOSED to reverse voltages and OPEN to forward voltages. Semiconductor diode and idealized diode characteristics are shown in Fig. 14.9a and b. The diode model of Fig. 14.9c with $V_F = 0.7$ V is suitable for the analysis of logic circuits using silicon diodes. For germanium



(a) Semiconductor diode (b) Idealized diode (c) Circuit model (d) Diode **OR** gate

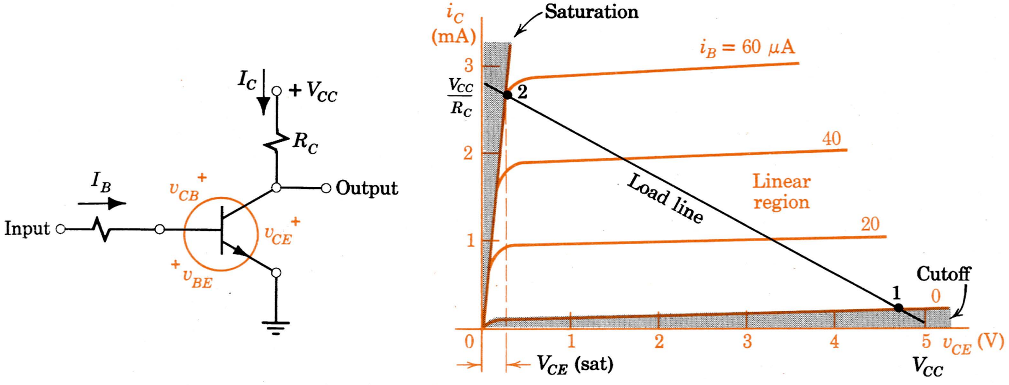
Figure 14.9 The diode as a switch.

diodes, $V_F = 0.3$ V, and where I_s is significant, the more sophisticated model of Fig. 12.4b must be used.

The highly nonlinear characteristic of the diode isolates the several inputs of a logic gate. The output of an **OR** gate (Fig. 14.9d) always follows the most positive of the input signals; diodes with less positive inputs are reverse biased. After several successive **OR** gates, the output signal will be significantly degraded. Eventually it is necessary to restore the signal, and a transistor inverter can be used for this purpose. In alternating **AND** and **OR** gates, the diode voltage drops tend to cancel and the logic level is maintained. This idea is used in designing complicated diode logic circuits.

Transistor Switches

The performance of logic circuits can be improved by taking advantage of the switching capability of the transistor. As you recall, a transistor amplifier is operated in the *linear* or *normal* region with the emitter-base junction forward-biased and the collector-base junction reverse-biased. If both junctions are reverse-biased, however, practically no collector current flows and the transistor is said to be operating in the *cutoff region.* In the basic switching circuit of Fig. 14.10a, if the input voltage is zero, there is no base current and operation is at point 1 in Fig. 14.10b. The collector current is practically zero ($I_C \cong I_{CEO}$) and the *switch* whose contacts are the collector and emitter terminals is OPEN. The cutoff current is exaggerated in Fig. 14.10b; a typical value of collector current of less than 1 μA with an applied voltage of 5 V corresponds to a dc *cutoff resistance* of more than 5 MΩ.



(a) Switching circuit

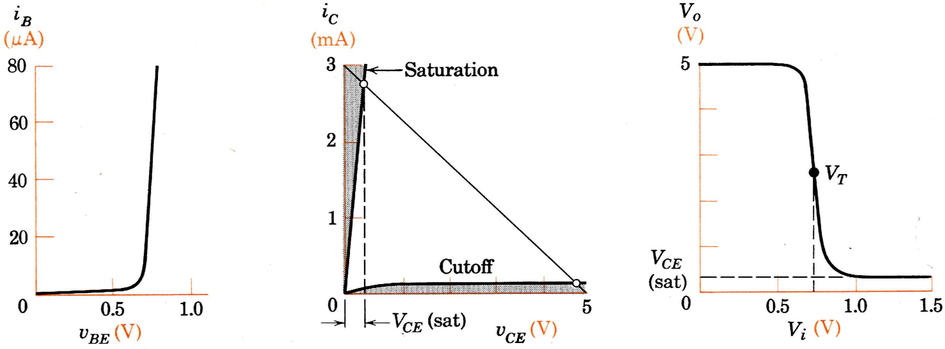
(b) Operating regions

Figure 14.10 The transistor as a switch.

A positive voltage pulse applied to the input terminal forward-biases the emitter-base junction, causes an appreciable base current, and moves operation to point 2. An increase in base current above 60 μA produces no further effect on collector current and the transistor is said to be operating in the *saturation region.* (This limit on collector current is represented in the dc transistor model of Fig. 14.12b by $I_C \leq \beta I_B$.) The voltage drop across the "switch" is called *collector saturation* voltage $V_{CE(\text{sat})}$ and is typically a few tenths of a volt. Note that if v_{CE} is less than v_{BE}, the collector-base junction is also forward-biased. As indicated in Fig. 14.10b, a collector current of around 3 mA at a saturation voltage of 0.3 V corresponds to a *saturation resistance* of around 100 Ω. When the transistor switch is CLOSED, the collector current is determined primarily by the load resistance, and $I_C \cong V_{CC}/R_C$.

Transistor Logic Gates

The transistor switch of Fig. 14.10 is basically an RTL inverter or **NOT** gate. When the input is near zero (logic **0**), the output is near +5 (logic **1**); a positive input (**1**) causes current flow and the output drops to $V_{CE(\text{sat})}$(**0**). The *transfer characteristic,* i.e., the relationship between input and output voltages, is shown in Fig. 14.11. The *threshold voltage* V_T is approximately 0.7 V.† Such switches can be connected in series or parallel to provide **NAND** or **NOR** gates.



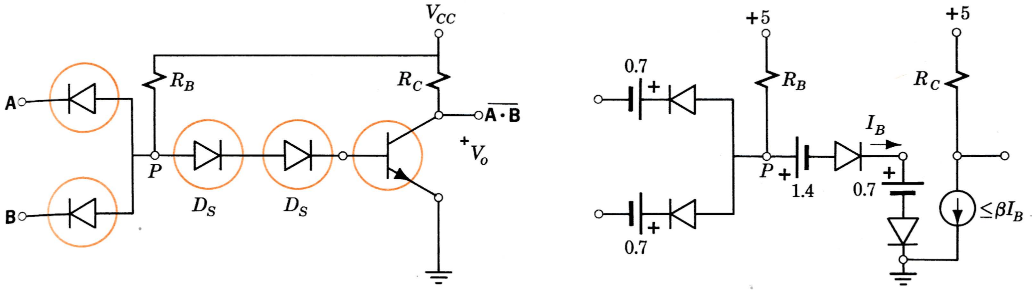
(a) Input (b) Output (c) Transfer

Figure 14.11 Characteristics of an *RTL* switch.

A more sophisticated, and faster, device is the DTL **NAND** gate of Fig. 14.12. It is faster because the charging currents accompanying a change in transistor state flow through the low forward resistances of diodes instead of through the higher series resistances necessary in RTL operation. A fast-acting Schottky diode may be added in parallel with the collector-base junction to limit the forward

---

† In practice, the circuit designer may hold v_{BE} below 0.4 V, say, to ensure cutoff and provide $v_{BE} = 0.8$ V to ensure saturation.



(a) Circuit diagram

(b) DC model of input circuit

Figure 14.12 Operation of a *DTL* NAND gate.

bias and reduce the turn-off delay. The greater sophistication is inexpensively realized in *IC* form.

If this gate follows other DTL switches in the OFF state, inputs A and B are approximately 5 V (1), and diodes A and B are reverse biased. Series diodes D_S are forward biased by V_{CC} and the high base current holds the transistor in saturation, so the output is about 0.3 V (0). The potential at point P is about

Example 2

For the DTL NAND gate of Fig. 14.12a, $V_{CC} = 5$ V, and $R_B = R_C = 5$ kΩ. For the transistor, $\beta = 30$ and $V_{CE(\text{sat})} = 0.3$ V. If input A is 5 V and input B is 0.3 V, determine the no-load output voltage V_o. NOTE: In dealing with nonlinear "threshold" devices, circuit equations involve inequalities. The practical approach is to assume a state (ON, OFF) and check to see if that state is consistent with the data.

For input $A = 5$ V and $B = 2$ V, determine the base current, the collector current, and the no-load output voltage.

The lowest input governs the state of a NAND gate. For $B = 0.3$ V, V_P cannot exceed $0.3 + 0.7 = 1.0$ V. If current did flow through the series diodes,

$$V_{BE} = V_P - 2(0.7) = 1.0 - 1.4 = -0.4 \text{ V}$$

But a forward bias of $V_{BE} \geq +0.7$ V is required, ∴ no base current flows, the transistor is cut off, and

$$V_o = V_{CC} - I_C R_C = 5 - 0 = 5 \text{ V}.$$

For $B = 2$ V, V_P cannot exceed $2 + 0.7 = 2.7$ V. The critical value for turn on is $V_P = 3(0.7) = 2.1$ V. The lowest value governs, ∴ base current flows and

$$I_B \cong (V_{CC} - 2.1)/R_B = (5 - 2.1)/5\text{k} = 0.58 \text{ mA}$$

But the largest possible no-load collector current is

$$I_C \cong V_{CC}/R_C = 5/5\text{k} = 1 \text{ mA}$$

which requires a base current of only

$$I_B = I_C/\beta = 1/30 \cong 0.033 \text{ mA}$$

∴ the transistor is in saturation and

$$V_o = V_{CE(\text{sat})} = 0.3 \text{ V}$$

$2 \times 0.7 + 0.7 = 2.1$ V. If now input A drops, at $V_T = 2.1 - 0.7 = 1.4$ V, diode A is forward biased and begins to conduct. The potential at point P drops, the transistor is cut off, and V_o rises to $+5$ V (**1**). Note that a change in input from 5 to 1.4 or 3.6 V is required to switch from the high state, and a change from 0.3 to 1.4 or 1.1 V is required to switch from the low state. The difference between the operating input voltage and the threshold voltage is called the *noise margin* since unwanted voltages lower than this value will not cause false switching. (See Example 2 p. 373.)

Typically, the output of the **NAND** gate of Example 2 is connected to the inputs of other DTL gates. When V_o is high, the next input diodes are reverse biased and no "load" current is drawn. When V_o is low, however, each input diode draws current equal to $(V_{CC} - V_P) \div R_B = (5 - 0.7 - 0.3) \div 5000 = 0.8$ mA in this example. The base current I_B must put the transistor into saturation and also supply additional collector current ($I_C = \beta I_B$) to "drive" the load represented by following gates. The maximum fan-out for DTL gates is about 8. (See Problem 4.)

As another example of the sophisticated gates available as integrated circuits, consider the TTL gate in Fig. 14.13 where T_1 is a multiple-emitter transistor

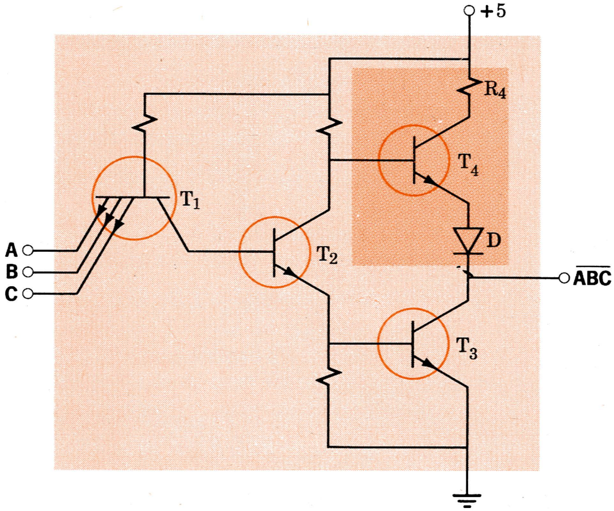


Figure 14.13 An integrated *TTL* **NAND** gate.

that acts as an **AND** gate. If **A AND B AND C** are **1** (+5 V), the emitter junctions of T_1 are reverse biased, but the base-collector junction is forward-biased providing the base current required by T_2 (ON) and T_3 (ON). With T_2 ON, the base-emitter potential of T_4 is too low for forward bias and T_4 is OFF. In other words, if all inputs are **1**, T_2 is ON, T_3 is ON (T_4 is OFF), and the output of this **NAND** gate is **0**. Diode D adds a forward voltage drop to ensure that T_4 is OFF when T_3 is ON.

If one or more inputs becomes 0 (+0.3 V), the emitter of T_1 is effectively grounded, T_1 is forward biased, and its collector potential drops turning OFF T_2 and T_3 (and turning ON T_4 since the collector potential of T_2 rises), and the output becomes I. In effect the combination of T_4 and R_4 is a variable load resistance for T_3, a high resistance when T_3 is conducting so that power consumption is low and a low resistance when T_3 turns OFF for fast switching. Since all capacitances associated with the inputs to the next logic level must be charged to nearly V_{CC} when T_3 turns OFF, the delay is reduced if the charging current is supplied through a low value of R_4. In large-volume production, the fabrication of such an integrated circuit is little more expensive than fabrication of individual transistors.

MOS Logic Gates

The *p*-channel enhancement-mode MOS transistor has several virtues that make it useful in digital circuits; it requires only a small area, it is normally OFF, and it has a high input resistance. Furthermore, the basic gates can be fabricated from interconnected MOS structures only and therefore they are easy to produce in integrated circuit form. MOS logic is attractive in applications where the slow switching speed (due to the relatively large capacitances) is acceptable.

The characteristics in Fig. 14.14 indicate the possibility of using a MOSFET as a binary switch or inverter. Note that the supply voltage is higher than that for the BJT, and the "low" output voltage is considerably higher. Assuming a turn-on voltage of −4 V, reasonable logic levels might be: 0 to −3 V and −10 to −15 V. If the output of one inverter (v_{DS}) drives the input of another (v_{GS}), any value of v_{GS} between 0 and −3 V would be insufficient for turn-on and any value between −10 and −15 V would provide a strong turn-on. In practice, the voltage state near zero is assigned a logic 0 and the large negative voltage is assigned a logic I. (This is called "negative logic" in contrast to the "positive logic" used with BJTs.)

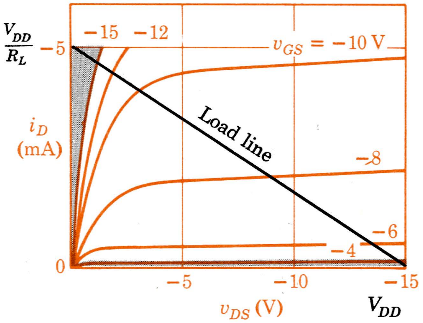
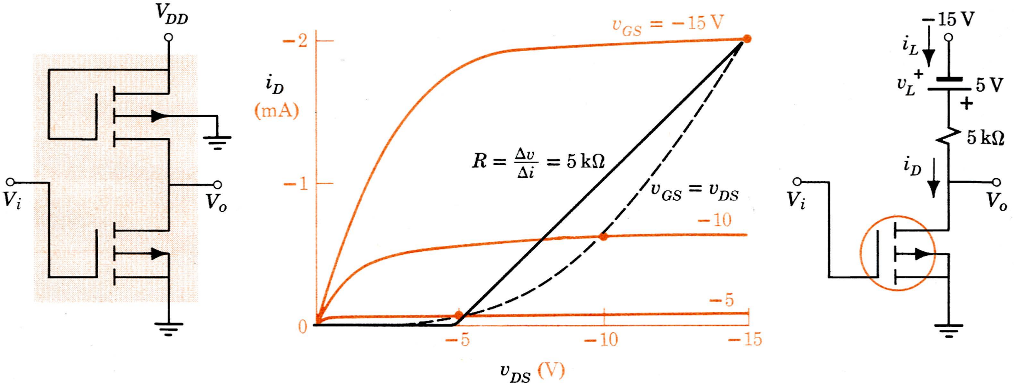


Figure 14.14 The MOSFET as a switch.

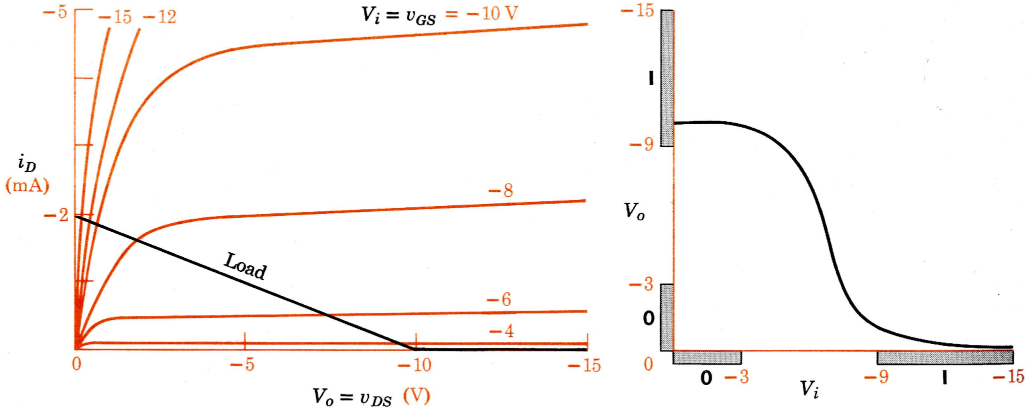
In the inverter of Fig. 14.15, one MOS device is used as a *driver* and the other



(a) IC circuit (b) Load characteristic (c) Equivalent circuit

Figure 14.15 An inverter using only MOS devices.

as a passive *load.* With the gate of the load tied to the drain, $v_{GS} = v_{DS}$ and the i_L-v_L characteristic is the dashed line in Fig. 14.15b. The load can be modeled, approximately, by a 5-V source in series with a 5-kΩ resistance (the channel resistance for the load is much higher than that for the driver), and the equivalent circuit is as shown. Note that output voltage $V_o = V_{DD}$-v_L. When the load characteristic is reversed and superimposed on the driver characteristics (Fig. 14.16a), the **NOT** function is revealed. By using multiple drivers, multiple-input **NAND** or **NOR** gates are obtained. (See Exercise 18.)



(a) Load and driver (b) Transfer characteristic

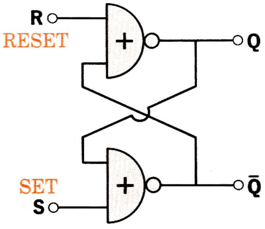
Figure 14.16 Operation of an MOS inverter.

Memory Devices

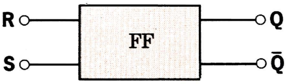
The outputs of the basic logic gates are determined by the present inputs; in response to the various inputs, these *combinational* circuits make "decisions." Along with these *decision* components we need *memory* components to store instructions and results; the outputs of such *sequential* circuits are affected by past inputs as well as present.

What characteristics are essential in a memory unit? A binary storage device must have two distinct states, and it must remain in one state until instructed to change. It must change rapidly from one state to the other, and the state value (0 or 1) must be clearly evident. The *bistable multivibrator* or *flip-flop*, a simple device that meets these requirements inexpensively and reliably, is used in all types of digital data processing systems.

A Logic Gate Memory Unit



(a) **NOR** gate realization



(b) RS flip-flop symbol

Figure 14.17 The flip-flop, a basic memory unit.

First let us analyze a basic memory unit consisting of familiar logic gates. In the NOR gate flip-flop of Fig. 14.17, the output of each NOR gate is fed back into the input of the other gate. The operation is summarized in Table 14-1 where we assume to start that Q_0, the *present* state of the output Q, is 0 and inputs to the *set* terminal S and the *reset* (or *clear*) terminal R are both 0.

To SET the flip-flop, a 1 is applied to S only. For $Q_0 = 0$, $\bar{Q} = \overline{Q_0 + S} = 0$, and $Q = \overline{R + \bar{Q}} = 1$, the output is inconsistent with the input, the system is *unstable*, and Q must *flip*. After Q changes, Q_0 (the present state Q) changes to 1, and $\bar{Q}$ becomes $\overline{1 + 1} = 0$; hence $Q = \overline{0 + 0} = 1$, a *stable* state. (In this analysis, keep in mind that *if either input to a* NOR *gate is* 1, *the output is* 0.) Removing the input from S causes no change. We conclude that Q = 1, and $\bar{Q} = 0$ is the stable state after being SET. Applying another input to S produces no change.

Table 14-1 Analysis of a Memory Unit

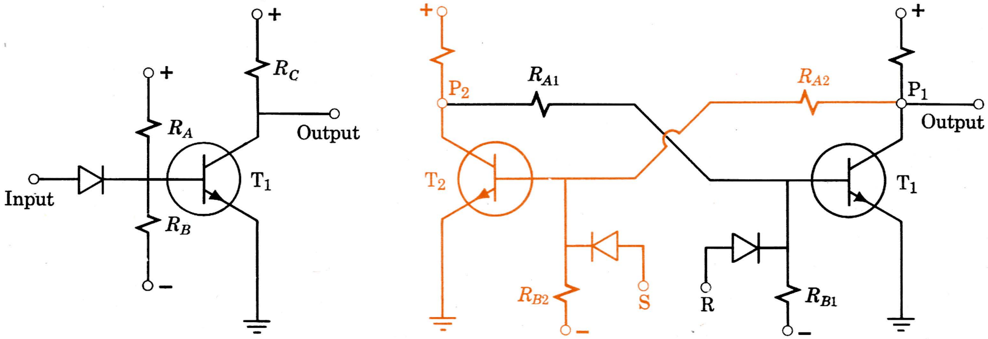
Action	Q_0	S	R	$\bar{Q}$	Q	Conclusion
Assume	0	0	0	1	0	This is a stable state.
Apply 1 to S	0	1	0	0	1	Unstable state; Q changes.
(Q_0 becomes 1)	1	1	0	0	1	Stable.
Remove 1 from S	1	0	0	0	1	The stable state after SET.
Apply 1 to S again	1	1	0	0	1	No change in Q.
Remove 1 from S	1	0	0	0	1	The stable state after SET.
Apply 1 to R	1	0	1	0	0	Unstable state; Q changes.
(Q_0 becomes 0)	0	0	1	1	0	Stable.
Remove 1 from R	0	0	0	1	0	The stable state after RESET.
Apply 1 to S and R	0	1	1	0	0	Unacceptable; $Q = \bar{Q}$.

To RESET the flip-flop, a I is applied to R only. This results in an unstable system and Q must *flop* to 0. (You should perform the detailed analysis.) A change in Q_0 to 0 produces a stable output Q = 0. Removing the input to R or applying another input to R produces no change. We conclude that Q = 0 and $\bar{Q}$ = I is the stable state after being RESET.

Note that only a momentary input is required to produce a complete transition; this means that very short pulses can be used for triggering. Attempting to SET and RESET simultaneously creates an undesirable state with both Q and $\bar{Q}$ = 0. This state ambiguity is unacceptable in a bistable unit and actual circuits are designed to avoid this condition.

A Transistor Flip-Flop

The operation of an electronic flip-flop is based on the switching and amplifying properties of a transistor. In the switch in Fig. 14.18a, the voltage divider



(a) Transitor switch

(b) Elementary flip-flop

Figure 14.18 A transistor *RS* flip-flop.

R_A-R_B reverse-biases the base-emitter junction and the transistor is in cutoff or the switch is OPEN; because of inversion, a positive voltage appears at the output terminal. If a positive signal is applied to the input, the base-emitter junction is forward biased, the switch is CLOSED, and the output voltage drops to zero (nearly).

An input signal making the + terminal of R_A more positive could also forward bias the base-emitter junction and drop the output voltage to zero. This switch is a form of NOR gate; a flip-flop can be created from two such switches.

To follow the operation of the flip-flop (Fig. 14.18b), assume that T_1 is conducting (CLOSED) and T_2 is cut off (OPEN). With T_1 conducting, the potential of point P_1 is zero and, in combination with the negative voltage applied to R_{B2}, this ensures that T_2 is cut off. With T_2 cut off, the potential of point P_2 is large and positive and this supplies the bias that ensures that T_1 is conducting and V_o is low. In logic terms, Q = 0; this is a stable state that we may designate as the 0 state of this binary memory element.

A positive pulse applied to RESET terminal R has no effect since T_1 is already conducting. However, a positive pulse at SET terminal S causes T_2 to begin conducting, the potential of P_2 drops, the forward bias on T_1 is reduced, the potential of P_1 rises, the forward bias on T_2 increases, the potential of P_2 drops further, T_2 goes into saturation, and T_1 is cut off. The output voltage is high, $Q = 1$, and this indicates another stable state that we may designate as the 1 state. If a flip-flop in the 1 state receives a positive pulse at R, transition proceeds in the opposite direction (since the device is symmetric) and the device is RESET to the 0 state. In a well-designed flip-flop, these changes in state take place in a fraction of a microsecond, and we see that this simple device satisfies all the requirements of a binary storage element.

The T Flip-Flop

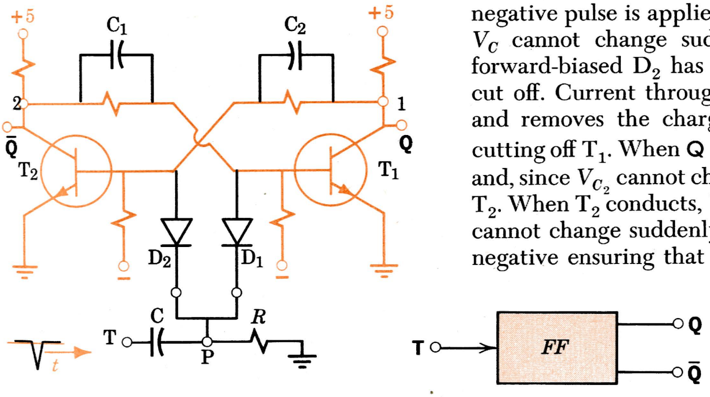
Practical flip-flops are available in a variety of forms. In the simple *trigger* or *toggle* flip-flop (see Example 3), each input pulse causes an abrupt change in the output state. The electronics can be designed for either positive or negative pulses, introduced into either base or collector circuits. By means of an *RC* differentiating circuit and rectifying diodes, the flip-flop can be made to trigger on the leading or trailing edge of a square wave. (See Exercise 23.)

Example 3

Modify a transistor RS flip-flop so that every time a negative pulse is applied to the *trigger* terminal T, the flip-flop changes state. (Note: A *negative* pulse can be used to *drop* the base potential and turn a conducting transistor OFF.)

For negative triggering, diodes D_1 and D_2 in Fig. 14.19 are reversed. To make sure the transition is completed, "speed-up" capacitors C_1 and C_2 are added. The negative trigger pulse is coupled in by capacitor C.

Assuming $Q = 0$ to start, T_1 is conducting, V_1 (at node 1) is near 0 V, and V_2 is near +5 V. When the negative pulse is applied at T, V_P drops sharply since V_C cannot change suddenly. Current through the forward-biased D_2 has no effect since T_2 is already cut off. Current through D_1 drops the base voltage, and removes the charge stored in the base of T_1, cutting off T_1. When Q changes to 1, V_1 rises abruptly and, since V_{C_2} cannot change suddenly, forward biases T_2. When T_2 conducts, V_2 drops and, since $V_{C_1}(= 5\text{V})$ cannot change suddenly, the base of T_1 goes sharply negative ensuring that T_1 is cut off.



(a) Modified RS circuit

(b) Symbol

Q_0	T	Q	$\bar{Q}$
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1
0	0	0	1

Figure 14.19 A *T* flip-flop.

The T flip-flop can be used in counting since each trigger pulse advances the Q output. A series of such flip-flops constitutes a *counter* (Fig. 14.20); the total number of input pulses is represented by the state of the flip-flops. Assuming all flip-flops have been RESET by the CLEAR pulse, the first COUNT triggers flip-flop A and registers $Q_A = 1$. At the next COUNT, $Q_A = 0$ and $\bar{Q}_A = 1$, supplying the trigger to flip-flop B which registers $Q_B = 1$; the total count is registered as 0010. Subsequent counts are shifted from one flip-flop to the next in the same way. Such a binary counter can also serve as a *frequency divider;* for a square wave input, the output of the fourth flip-flop will be a square wave of one-sixteenth the input frequency.

Q_D	Q_C	Q_B	Q_A	Count
0	0	0	0	0
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
				•
				•
				•
1	0	0	1	9

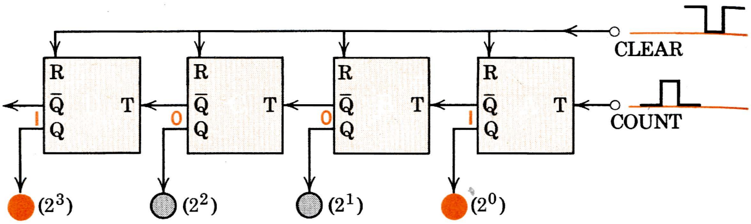
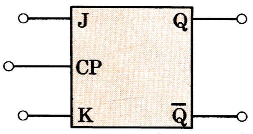


Figure 14.20 An elementary binary counter.

The JK Flip-Flop

A widely used memory element, available in integrated-circuit form, is the JK flip-flop. In the "clocked" type shown in Fig. 14.21, inputs at J and K are effective only when the clock pulse CP is 1; this permits timing the observation to coincide with the arrival of the desired inputs at J and K. Since the clock does the pulsing, the inputs need not be in pulse form. A JK flip-flop can be built from an RS unit by using NAND gates to control the inputs to R and S. (See Problem 6.)



(a) Symbol

J	K	Q
0	0	Q_0
0	1	0
1	0	1
1	1	$\bar{Q}_0$

(b) Truth table

Figure 14.21 A clocked *JK* flip-flop.

The output of a JK flip-flop depends on the states of the inputs at the time of the clock pulse. As indicated in the truth table, with 0 inputs at J and K, a clock pulse has no effect and the flip-flop remains in its present state Q_0. With different inputs, the unit behaves like an RS flip-flop. With $J = 1$ and $K = 0$, any clock pulse SETS the flip-flop to $Q = 1$; with $K = 1$ and $J = 0$, any clock pulse RESETS the flip-flop to $Q = 0$. (In other words, with $J \neq K$, $Q = J$.) With 1 inputs at both J and K (or, in the practical case, with J and K terminals unconnected), the unit behaves like a T flip-flop and toggles at each clock pulse.

In other versions of the basic JK package, DIRECT SET and DIRECT CLEAR terminals may be provided. The sophistication of the JK response makes it useful in a variety of digital computer applications such as counters, arithmetic units, and *registers*. Registers consisting of groups of identical flip-flops are used to "remember" or store binary information preparatory to processing.

As an illustration, consider the *serial shift register* used to accept binary data one bit at a time (Fig. 14.22). Here the flip-flop triggers on the *trailing edges* of a square pulse train (CP), corresponding to transitions from +5 to 0 V. The input number is introduced bit by bit into the J terminal of flip-flop A; 0110, for example, would be represented by logic levels 0, 1, 1, 0 at the four clock instants. The complement of the number is introduced into the K terminal. At CP 1, $J_A = 0$ and $K_A = 1$: therefore A is RESET and the first digit, 0, has been stored in flip-flop A. At CP 2, $J_B = 0$ and $K_B = 1$: therefore B is RESET and the first digit has been shifted to flip-flop B. Meanwhile, $J_A = 1$ and $K_A = 0$: therefore A is SET and the second digit has been stored in flip-flop A. After four such clock pulses, the number 0110 has been entered into the register and is available (as indicated by the "lights") for processing. A digital computer is just a vast assemblage of simple logic units organized into input/output units, control systems, memory sections, and processing systems.

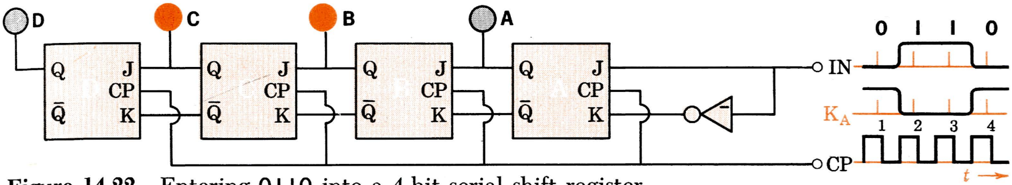
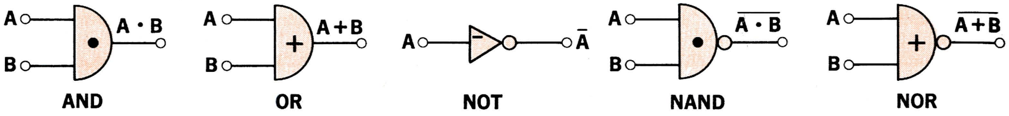


Figure 14.22 Entering 0110 into a 4-bit serial shift register.

SUMMARY

- Digital logic circuits employ discrete instead of continuous signals.
The basic logic operations are: **AND**, **OR**, **NOT**, **NAND**, and **NOR**.



All operations can be achieved using only **NAND** gates or only **NOR** gates.
The validity of any logic statement can be demonstrated with a truth table.

- Logic circuits are classified in terms of the components (R, D, T) employed.
Design factors include: cost, noise immunity, power consumption, signal degradation, fan-in, fan-out, speed, and reliability.

- Diodes and transistors are used as switches in electronic logic gates.
Transistors are switched from cutoff to saturation by changes in i_B.
Integrated circuits permit high-performance gates at reasonable cost.

- Memory components are needed to store data processed by decision components.
A flip-flop must change rapidly from one distinct stable state to the other.
The transistor flip-flop is an inexpensive, widely used binary storage device.

- The basic RS flip-flop is SET by $S = 1$ and RESET by $R = 1$.
The T flip-flop is toggled from one state to the other by successive triggers.
The JK flip-flop operates on each clock pulse, as determined by the inputs.
Basic applications of flip-flops include counters and registers.

REFERENCES

1. A. P. Malvino and Donald P. Leach, *Digital Principles and Applications,* McGraw-Hill Book Co., New York, 1969.
An excellent introduction to digital systems with detailed discussions and many illustrative examples. Includes basic concepts and their practical applications in computers.
2. Robert G. Hibberd, *Integrated Circuits: A Basic Course for Engineers and Technicians,* McGraw-Hill Book Co., New York, 1969.
A clear qualitative discussion of logic circuits and their integrated circuit realization. Compares the properties of the various IC logic families.
3. John B. Peatman, *The Design of Digital Systems,* McGraw-Hill Book Co., New York, 1972.
A well-written introduction to modern digital devices and systems and their engineering applications.
4. J. N. Harris, P. E. Gray, and C. L. Searle, *Digital Transistor Circuits* (*SEEC,* Vol. 6), John Wiley & Sons, New York, 1966.
An outstanding presentation of fundamental theory and design principles with emphasis on the electronic aspects of digital circuits. Quantitative treatment of switching speed and other circuit parameters.

REVIEW QUESTIONS

1. Distinguish between analog and digital signals; between gates and switches.
2. Why did the successful computer follow the invention of the transistor?
3. Explain how diodes and transistors function as controlled switches.
4. What is a "truth table"? How is it used?
5. Draw symbols for and distinguish between **AND**, **OR**, **NOR**, and **NAND** operations.
6. What is "**NOT** the complement of variable A inverted?"
7. Draw a **NAND** gate. Why are they used so widely?
8. How are electronic logic circuits classified? What is RTL? DTL? TTL?
9. List the factors of importance in the practical design of logic circuits.
10. Show how alternating **AND** and **OR** gates preserves the logic level.
11. Show with a sketch how cutoff and saturation resistances are calculated.
12. Draw a DTL **NOR** gate and explain its operation.
13. Distinguish between decision and memory components; give an example of each.
14. What is a flip-flop? What is its function in a computer?
15. Explain the operation of a flip-flop consisting of two **NOR** gates.
16. Explain the operation of a transistor RS flip-flop.
17. Distinguish between RS, T, and JK flip-flops.
18. Explain how T flip-flops "count." How many are needed to count to ten?
19. Explain the purpose and operation of a serial shift register.

EXERCISES

1. Devise a two-input **NAND** gate using relays (Fig. 14.1) and construct the truth table.
2. Using relays (Fig. 14.1), devise an **Equality** gate ($f = A \cdot B + \bar{A} \cdot \bar{B}$) and construct the truth table. (Hint: double-pole, double-throw, relay switches are available.)
3. The "resistor logic" circuit of Fig. 14.23 has binary inputs of 0 and 4 V.

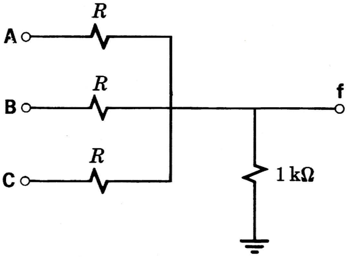


Figure 14.23

(a) For $R = 1\ \text{k}\Omega$, construct a "truth table" showing actual voltages. If a detector with an adjustable threshold were available, what logic operations could be performed with this circuit?

(b) At what threshold should the detector be set for **AND** operation?

4. Replace the resistors R in Fig. 14.23 by diodes to form a three-input diode **OR** gate. Construct the truth table.
5. Repeat Exercise 4 to form an **AND** gate operating with negative inputs, i.e., **1** is represented by -5 V and **0** by 0 V.
6. How many rows are needed in the truth table of an n-input logic circuit?
7. Assuming 5-V inputs and a diode forward voltage drop of 0.7 V, construct a "truth table" showing actual voltages for the diode circuits of (a) Fig. 14.3 and (b) Fig. 14.4.
8. When the transistor of Fig. 14.6b is conducting, $V_{BE} = 0.8$ V and $V_{CE} = 0.3$ V; $R = R_B = R_C = 2\ \text{k}\Omega$ and $I_{CEO} = 10\ \mu\text{A}$. The forward voltage drop on the diodes is 0.7 V. After several logic operations, the inputs at A are 1 and 2 V. Assuming 0 input at B, estimate the output voltages.
9. The transistor and diodes of Exercise 8 are used in Fig. 14.7b with $R_C = 2\ \text{k}\Omega$ and $R_B = 20\ \text{k}\Omega$. Assuming a 5-V input at B, estimate the output for inputs of 1 and 2 V at A.

10. A portion of a table defining the performance of the basic gates is given:

Gate	Inputs	Output
AND	All inputs **1**	**1**
	Any one input **0**	**0**

Complete the table for **NAND**, **OR**, **NOR**, and **NOT** gates.

11. Use **NOR** gates to form a two-input **AND** gate and construct the truth table as a check.

12. For a 2N3114 transistor switch (Fig. A3) operating at $V_{CC} = 5$ V with $R_C = 100\ \Omega$, estimate the typical ON and OFF resistances at room temperature. Repeat for the worst possible conditions at 150°C.

13. (a) Classify the type of logic represented in Fig. 14.24 and identify the logic operation performed. On a sketch of the transistor characteristics, label the points corresponding to voltages ($A = 0$, $B = 0$) and ($A = 0$, $B = 5$).

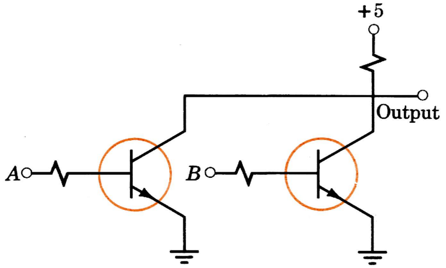


Figure 14.24

(b) Devise an **AND** gate using the same type of logic.

14. In the silicon **NAND** gate of Fig. 14.12, $R_B = 2\ \text{k}\Omega$, $R_C = 5\ \text{k}\Omega$, and $\beta \cong 50$. The "load" across which V_o appears is 15 kΩ.

(a) For $V_A = 0.2$ V and $V_B = 0.2$ V, predict V_o.

(b) For $V_A = 3$ V and $V_B = 4$ V, predict V_o.

15. For the **NAND** gate of Fig. 14.7b, $R_B = 2\ \text{k}\Omega$, $R_C = 4.7\ \text{k}\Omega$, $\beta = 50$, and the current drawn at the output is negligible.

(a) Sketch the i_B vs v_{BE} characteristic for the transistor and estimate the critical values of v_{BE} for the ON and OFF states.

(b) Sketch the i_C vs v_{CE} characteristics and estimate the critical values of i_C and output voltage V_o.

(c) Replace the active devices by dc circuit models and redraw the circuit. Assuming V_B is "high," estimate the critical values of V_A for V_o "high" and "low," and sketch the transfer characteristic V_o vs V_A.

16. An RTL 3-input **NOR** gate uses the transistor of Fig. 14.10 with $V_{CC} = 5$ V and $R_C = 5000\ \Omega$.

(a) Draw the circuit with input resistors $R_a = R_b = R_c = 3000\ \Omega$.

(b) For $V_A = 0$ V, $V_B = 1$ V, and $V_C = 2$ V, predict V_o.

17. Assume T_3 in Fig. 14.13 is ON and estimate the voltage at the base of each transistor. Prove that T_4 is OFF.

18. Design, i.e. draw the circuit diagram for, a 2-input **NAND** gate using MOS devices only.

19. Repeat Exercise 18 for a 2-input **NOR** gate.

20. Design a basic memory unit consisting of 2 **NAND** gates and describe its operation in a table similar to Table 14-1.

21. Draw the circuit diagram for a flip-flop using MOS devices only.

22. For the transistors of Fig. 14.18, $V_{CE(\text{sat})} = 0.3$ V and $V_{CC} = 5$ V. Draw up a table showing the voltages at P_1 and P_2 for the following conditions: (a) T_1 initially conducting, (b) a negative pulse applied to S, (c) a positive pulse applied to S, (d) a negative pulse applied to R, (e) a positive pulse applied to R.

23. The circuit of Fig. 14.25 is used to develop trigger pulses for a flip-flop with a relatively high input impedance.

(a) For a square-wave signal, sketch the input voltage to the flip-flop.

(b) Repeat part (a) when a diode is inserted with easy current flow from *a* to *b*.

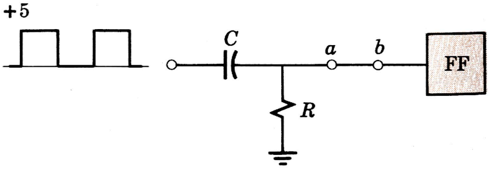


Figure 14.25

(c) For $R = 10\ \text{k}\Omega$ and a square-wave frequency of 500 Hz, design a circuit to provide only negative pulses with time constants of $20\ \mu\text{s}$.

24. The square wave in Fig. 14.25 has a frequency of 500 Hz. The T flip-flop is triggered on negative pulses as in Fig. 14.19.
(a) Draw 6 cycles of the square wave along with the corresponding flip-flop output.
(b) Add to part (a) the output of a second trigger circuit and T flip-flop connected in tandem.
(c) What operation is performed by the combination of flip-flops?

25. A CMOS digital device is fabricated as shown in Fig. 14.26. Draw a clearly labeled circuit diagram using standard symbols. If $|V_T|$ is 2.3 V for each unit, determine the outputs for inputs of 1 and 9 V. What logic function is performed?

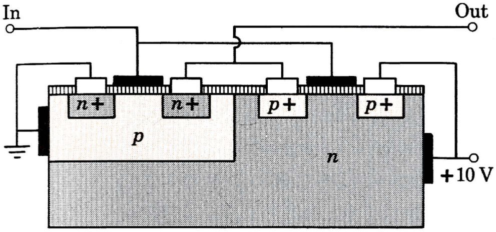


Figure 14.26

26. Analyze the memory element in Fig. 14.27. Draw up a truth table assuming the input is a series of pulses and Q is initially 0. What function is performed by the AND gates?

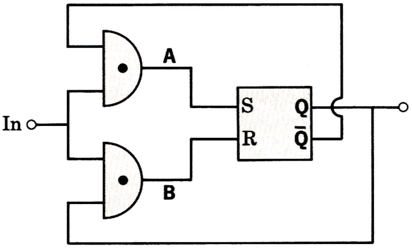


Figure 14.27

27. Two JK flip-flops which respond to downward transitions are connected in tandem (Fig. 14.28 with J = K = 1). For a 2-kHz square-wave input, determine the output.

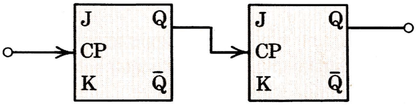


Figure 14.28

28. A JK flip-flop which responds to downward transitions is connected as in Fig. 14.29. For a 2-kHz square-wave input, determine the output.

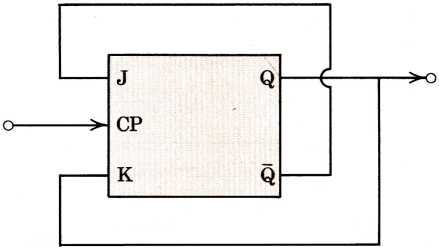


Figure 14.29

29. Design (i.e., draw a schematic diagram of) a binary counter that can count up to decimal 1000.

PROBLEMS

1. Devise a dc model for use in the analysis of a 2N3114 transistor (Fig. A3) as a switch: (a) neglecting and (b) including the effect of collector cutoff current.
2. Design (i.e., sketch the construction of) an integrated circuit realization of the DTL NAND gate in Fig. 14.7b following the form of Fig. 13.24.
3. The RTL NOR gate of Exercise 16 feeds the *A* terminals of three similar RTL gates (a fan out of 3).
   (a) Draw the circuit showing the connections

to input resistors R_{A1}, R_{A2}, and R_{A3}.

(b) For $V_A = V_B = V_C = 0$ for this NOR gate, predict V_o.

(c) Estimate the low-state noise margin for a fan-out of 3.

4. A four-input DTL NAND gate similar to Fig. 14.12a "drives" eight similar NAND gates (a fan-out of 8). Assume $V_{CC} = 5$ V, $R_B = R_C = 5$ kΩ ± 10%, $\beta = 30 \pm 40\%$, $V_{CE(\text{sat})} = 0.3$ V, and $V_{\text{junct}} = 0.7 \pm 0.1$ V. Draw the circuit, define the "worst possible" operating condition, and predict the "low" output voltage under this condition.

5. In the ECL gate of Fig. 14.30, $R_C = R_3 = 2$ kΩ and $R_E = R_4 = 5$ kΩ.

(a) For $V_A = V_B = 2.9$ V, estimate the voltages at *P*, *Q*, and Output.

(b) Repeat part (a) for $V_A = V_B = 4.4$ V.

(c) Repeat part (a) for $V_A = 2.9$ V and $V_B = 4.4$ V and identify the function performed.

6. Starting with an available RS flip-flop, use basic logic gates to obtain an elementary JK flip-flop. Starting with the truth table for the RS element, build up the overall truth table.

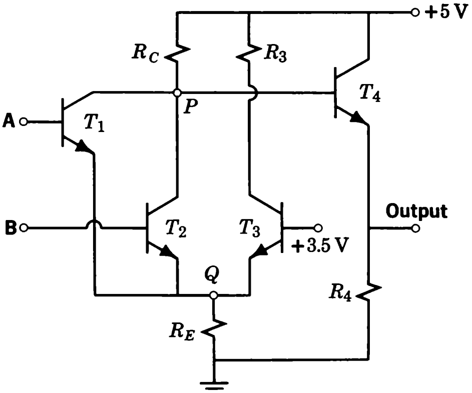


Figure 14.30 ECL gate.

7. Four T flip-flops are connected in cascade to form a scale-of-sixteen counter. Draw a schematic diagram indicating the paths taken by pulses. Using feedback links to reintroduce counts into earlier stages, convert the unit into a decimal scalar, i.e., a counter with one output pulse for every ten input pulses.

Chapter 15

Logic Circuits

Binary Number System
Boolean Algebra
Combinational Circuits
Computer Organization

Digital information processing requires special electronic components. In Chapter 14 we saw how diodes and transistors can be used as switches in gates and flip-flops. The efficient design of digital circuits also requires a special numbering system and a special algebra. Our purpose here is to examine the binary number system and learn to apply logic theorems to binary relations. With this background we can analyze given logic circuits or synthesize new ones to realize desired logic functions. Finally, we look at the organization of a modern digital computer to see how combinational and sequential logic circuits are used in high-speed computation.

Binary Number System

In the decimal system a quantity is represented by the *value* and the *position* of a digit. The number 503.14 means

$$500 \qquad + 0 \qquad + 3 \qquad + \tfrac{1}{10} \qquad + \tfrac{4}{100}$$

or

$$5 \times 10^2 + 0 \times 10^1 + 3 \times 10^0 + 1 \times 10^{-1} + 4 \times 10^{-2}$$

In other words, 10 is the *base* and each position to the left or right of the decimal point corresponds to a *power* of 10. Perhaps it is unfortunate that we do not have 12 fingers, because in certain ways 12 would be a better base. In fact, such a base-twelve or *duodecimal* system was used by the Babylonians, and we still use 12 in subdividing the foot, the year, and the clock face.

In representing data by ON-OFF switch position, there are only two possibilities and the corresponding numbers are **1** and **0**. In such a *binary* system the base is 2 and the total number of fingers on both hands is written **1010** since

$$1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 8 + 0 + 2 + 0 = 10$$

In electronic logic circuits the numbers **1** and **0** usually correspond to two easily distinguished voltage levels specified by the circuit designer. For example, in DTL and TTL, **0** corresponds to a voltage near zero and **1** to a voltage of approximately +5 V.

Number Conversion

Binary-to-Decimal-Conversion. In a binary number, each position to the right or left of the "binary point" corresponds to a power of 2, and each power of two has a decimal equivalent.

To convert a binary number to its decimal equivalent, add the decimal equivalents of each position occupied by a **1**.

For example,

$$\begin{aligned} \mathbf{110001} &= 2^5 + 2^4 + 0 + 0 + 0 + 2^0 = 32 + 16 + 1 = 49 \\ \mathbf{101.01} &= 2^2 + 0 + 2^0 + 0 + 2^{-2} = 4 + 1 + \tfrac{1}{4} = 5.25 \end{aligned}$$

Decimal-to-Binary Conversion. A decimal number can be converted into its binary equivalent by the inverse process, i.e., by expressing the decimal number as a sum of powers of 2. An automatic, and more popular, method is the *double-dabble* process in which integers and decimals are handled separately.

To convert a decimal integer to its binary equivalent, progressively divide the decimal number by 2, noting the remainders; the remainders taken in reverse order form the binary equivalent.

Example 1

Convert decimal 28.375 into its binary equivalent.

Using the double-dabble method on the integer, (a shorthand notation is shown at the left)

2 | 28		
14	0	$28 \div 2 = 14$ with a remainder of 0
7	0	$14 \div 2 = 7$ with a remainder of 0
3	1	$7 \div 2 = 3$ with a remainder of 1
1	1	$3 \div 2 = 1$ with a remainder of 1
0	1	$1 \div 2 = 0$ with a remainder of 1

The binary equivalent is **11100**.

Then converting the fraction,

$$\begin{aligned} 0.375 \times 2 &= 0.75 \text{ with a carry of } 0 \\ 0.75 \times 2 &= 1.50 \text{ with a carry of } 1 \\ 0.50 \times 2 &= 1.00 \text{ with a carry of } 1 \end{aligned}$$

The binary equivalent is **.011**.

28.375 is equivalent to binary **11100.011**

To convert a decimal fraction to its binary equivalent, progressively multiply the fraction by 2, removing and noting the carries; the carries taken in forward order form the binary equivalent. (See Example 1.)

Binary Arithmetic

Since the binary system uses the same concept of value and position of the digits as the decimal system, we expect the associated arithmetic to be similar but easier. (The binary multiplication table is very short.) In *addition*, we add column by column, carrying where necessary into higher position columns. In *subtraction*, we subtract column by column, borrowing where necessary from higher position columns. In subtracting a larger number from a smaller, we can subtract the smaller from the larger and change the sign just as we do with decimals.

Example 2

Convert the given numbers to the other form and perform the indicated operations.

$$\begin{array}{r} 14 \\ +11 \\ \hline 25 \end{array} \quad \begin{array}{r} 1110 \\ +1011 \\ \hline 11001 \end{array} \qquad \begin{array}{r} 13 \\ -10 \\ \hline 3 \end{array} \quad \begin{array}{r} 1101 \\ 1010 \\ \hline 0011 \end{array} \qquad \begin{array}{r} 1010 \\ -1101 \\ \hline -0011 \end{array} \quad \begin{array}{r} 10 \\ -13 \\ \hline -3 \end{array}$$

In *multiplication*, we obtain partial products using the binary multiplication table ($0 \times 0 = 0, 0 \times 1 = 0, 1 \times 0 = 0, 1 \times 1 = 1$) and then add the partial products.† In *division*, we perform repeated subtractions just as in long division of decimals.

Example 3

Convert the given numbers to the other form and perform all the indicated operations.

(a)
$$\begin{array}{r} 14.5 \\ \times 1.25 \\ \hline 725 \\ 290 \\ 145 \\ \hline 18.125 \end{array}$$

(b)
$$\begin{array}{r} 101.1 \\ 11.1\overline{)10011.01} \\ 111 \\ \hline 1010 \\ 111 \\ \hline 111 \\ 111 \\ \hline 0 \end{array}$$

After decimal-binary conversion, the operations are:

(a)
$$\begin{array}{r} 1110.1 \\ \times 1.01 \\ \hline 11101 \\ 00000 \\ 11101 \\ \hline 10010.001 \end{array}$$

(b)
$$\begin{array}{r} 5.5 \\ 3.5\overline{)19.25} \\ 175 \\ \hline 175 \\ 175 \\ \hline 0 \end{array}$$

---

†Digital multiplying circuits are basically adding circuits.

Boolean Algebra

We see that binary arithmetic[†] and decimal arithmetic are similar in many respects. In working with logic relations in digital form, we need a set of rules for symbolic manipulation that will enable us to simplify complex expressions and solve for unknowns; in other words, we need a "digital algebra." Nearly 100 years before the first digital computer, George Boole, an English mathematician (1815–1864), had formulated a basic set of rules governing the true–false statements of logic. Eighty-five years later (1938), Claude Shannon (at that time a graduate student at MIT) pointed out the usefulness of *Boolean algebra* in solving telephone switching problems and established the analysis of such problems on a firm mathematical basis. From our standpoint, Boolean algebra is valuable in manipulating binary variables in **OR**, **AND**, or **NOT** relations and in the analysis and design of all types of digital systems.

Boole's Theorems

The basic postulates are displayed in Tables 15-1 and 15-2. At first glance, some of the relations in Table 15-1 are startling. Properly read and interpreted, however, their validity is obvious. For example, "**1 OR 1**" ($1 + 1$) is just equivalent to **1**, and "**0 AND 1**" ($0 \cdot 1$) is effectively **0**. In other words, for an **OR** gate with inputs of **1** at both terminals, the output is **1**, and for an **AND** gate with inputs of **0** and **1**, the output is **0**.

Table 15-1 Boolean Postulates in 0 and 1

OR	AND	NOT
$0 + 0 = 0$	$0 \cdot 0 = 0$	$\overline{0} = 1$
$0 + 1 = 1$	$0 \cdot 1 = 0$	$\overline{1} = 0$
$1 + 0 = 1$	$1 \cdot 0 = 0$	
$1 + 1 = 1$	$1 \cdot 1 = 1$	

Table 15-2 Boolean Theorems in One Variable

OR	AND	NOT
$A + 0 = A$	$A \cdot 0 = 0$	$\overline{\overline{A}} = A$
$A + 1 = 1$	$A \cdot 1 = A$	
$A + A = A$	$A \cdot A = A$	
$A + \overline{A} = 1$	$A \cdot \overline{A} = 0$	

In general, the inputs and outputs of logic circuits are *variables*, i.e., the signal may be present or absent (the statement is true or false) corresponding to the binary numbers **1** and **0**. The validity of the theorems in Table 15-2 may be

---

[†] Less circuitry is required in a digital computer if only adding circuits are used. To subtract a binary number, the computer first forms its *2's complement* by replacing each **0** or **1** by its complement and then adding **1**. Adding the 2's complement and dropping the last carry gives the same result as subtracting. For example (13 − 10),

$$1101 - 1010 \rightarrow 1101 + (0101 + 1) = 1101 + 0110 = 10011 \rightarrow 0011$$

reasoned out in terms of the corresponding logic circuit or demonstrated in a truth table showing all possible combinations of the input variables.

Example 4

Demonstrate the theorems

$$A + \bar{A} = 1 \quad \text{and} \quad A \cdot 1 = A$$

Using the postulates, the truth tables are:

A	$\bar{A}$	$A + \bar{A}$
0	1	1
1	0	1

A	$A \cdot 1$
0	0
1	1

Some of the more useful theorems in more than one variable are displayed in Table 15-3. The *commutation* rules indicate that the order of the variables in performing **OR** and **AND** operations is unimportant, just as in ordinary algebra. The *association* rules indicate that the order of the **OR** and **AND** operations is unimportant, just as in ordinary algebra. The second *distribution* rule indicates that variables or combinations of variables may be distributed in multiplication (not permitted in ordinary algebra) as well as in addition. The *absorption* rules are new and permit the elimination of redundant terms.

Table 15-3 Boolean Theorems in More Than One Variable

Commutation rules:
$A + B = B + A$
$A \cdot B = B \cdot A$

Association rules:
$A + (B + C) = (A + B) + C$
$A \cdot (B \cdot C) = (A \cdot B) \cdot C$

DeMorgan's theorems:
$\overline{A + B} = \bar{A} \cdot \bar{B}$
$\overline{A \cdot B} = \bar{A} + \bar{B}$

Absorption rules:
$A + (A \cdot B) = A$
$A \cdot (A + B) = A$

Distribution rules:
$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$
$A + (B \cdot C) = (A + B) \cdot (A + C)$

Example 5

Derive the absorption rule

$$A + (A \cdot B) = A$$

using other basic theorems.

A	$A \cdot B$	$A + (A \cdot B)$
0	0	0
1	B	1

Factoring by the second distribution rule,

$$A + (A \cdot B) = (A + A) \cdot (A + B) = A \cdot (A + B)$$

and we see that the two absorption forms are equivalent.

Substituting $A \cdot 1$ for $A \cdot A$ (Table 15-2),

$$A \cdot (A + B) = A \cdot (1 + B) = A \cdot 1 = A$$

The truth table is shown at the left.

DeMorgan's Theorems

Augustus DeMorgan, a contemporary of George Boole, contributed two interesting and very useful theorems. These concepts are easily interpreted in terms of logic circuits. The first says that a **NOR** gate ($\overline{A + B}$) is equivalent to an **AND** gate with **NOT** circuits in the inputs ($\bar{A} \cdot \bar{B}$). The second says that a **NAND** gate ($\overline{A \cdot B}$) is equivalent to an **OR** gate with **NOT** circuits in the inputs ($\bar{A} + \bar{B}$). As generalized by Shannon, DeMorgan's theorems say:

> *To obtain the inverse of any Boolean function, invert all variables and replace all* **OR***s by* **AND***s and all* **AND***s by* **OR***s.*

Example 6

Use DeMorgan's theorems to derive a combination of **NAND** gates equivalent to a two-input **OR** gate (Example 1, Chapter 14).

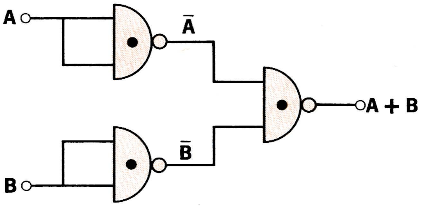


Figure 15.1 **OR** gate from **NAND** gates.

The desired function is $f = A + B$. Applying the general form of DeMorgan's theorems,

$$f = A + B = \overline{\bar{A} \cdot \bar{B}}$$

suggesting a **NAND** gate with **NOT** inputs. Since $\overline{A \cdot A} = \bar{A}$, a **NAND** gate with the inputs tied together performs the **NOT** operation. The logic circuit is shown in Fig. 15.1.

Logic Circuit Analysis

By comparing Example 6 with Example 1 in Chapter 14, we see that the theorems of Boolean algebra permit us to manipulate logic statements or functions directly, without setting up the truth tables. Furthermore, the use of Boolean algebra can lead to simpler logic statements that are easier to implement. This result is important when it is necessary to design a circuit to perform a specified logic function using the available gates—only **NAND** gates, for example. DeMorgan's theorems are particularly helpful in finding **NAND** operations that are equivalent to other operations.

Up to this point we have been careful to retain the specific **AND** sign in $A \cdot B$ to focus attention on the *logic* interpretation. Henceforth, we shall use the simpler equivalent forms AB and $A(B)$ whenever convenient.

The *analysis* of a logic circuit consists in writing a logic statement expressing the overall operation performed in the circuit. This can be done in a straight-forward manner, by starting at the input and tracing through the circuit, noting the function realized at each output. The resulting expression can be simplified or written in an alternative form using Boolean algebra. A truth table can then be constructed.

Example 7

Analyze the logic circuit of Fig. 15.2.

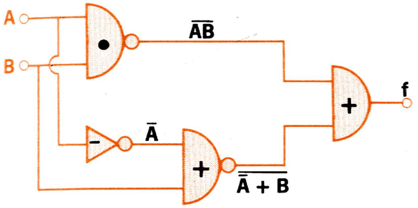


Figure 15.2 Logic circuit analysis.

Construct the truth table to demonstrate that this circuit could be replaced by a single **NAND** gate.

The sub-outputs are noted on the diagram. The overall function can be simplified as follows:

$$
\begin{aligned}
f &= \overline{AB} + \overline{\bar{A} + B} \\
&= (\bar{A} + \bar{B}) + A\bar{B} && \text{(DeMorgan's rule)} \\
&= \bar{A} + \bar{B}(1 + A) && \text{(Distribution)} \\
&= \bar{A} + \bar{B} && (1 + A = 1) \\
&= \overline{AB} && \text{(DeMorgan's rule)}
\end{aligned}
$$

A	B	$\overline{AB}$	$\overline{\bar{A} + B}$	f
0	0	1	0	1
0	1	1	0	1
1	0	1	1	1
1	1	0	0	0

Logic Circuit Synthesis

One of the fascinating aspects of digital electronics is the construction of circuits that can perform simple mental processes at superhuman speeds. A typical digital computer can perform thousands of additions of 10-place numbers per second. The logic designer starts with a logic statement or truth table, converts the logic function into a convenient form, and then realizes the desired function by means of standard or special logic elements.

The Half-Adder

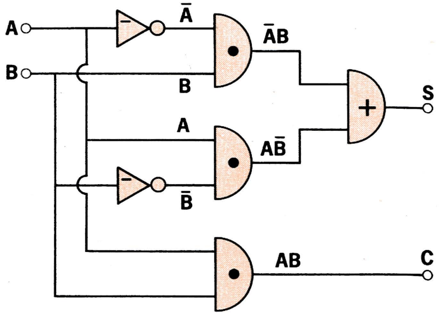
As an illustration, consider the process of addition. In adding two binary digits, the possible sums are as shown in Fig. 15.3a. Note that when $A = 1$ and $B = 1$, the *sum* in the first column is 0 and there is a *carry* of 1 to the next higher column. As indicated in the truth table, the half-adder must perform as follows: "S is 1 if A is 0 AND B is 1, OR if A is 1 AND B is 0; C is 1 if A AND B are 1." In logic nomenclature, this becomes

$$\begin{array}{rcccc} A = & 1 & 1 & 0 & 0 \\ +B = & 1 & 0 & 1 & 0 \\ \hline & 10 & 1 & 1 & 0 \end{array}$$

(a) Addition

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

(b) Truth table



(c) Half-adder circuit

Figure 15.3 Addition of two binary numbers.

$$S = \bar{A}B + A\bar{B} \quad \text{and} \quad C = AB \tag{15-1}$$

(A *full-adder* is capable of accepting the carry from the adjacent column.)

To *synthesize* a half-adder circuit, start with the outputs and work backward. Equation 15-1 indicates that the sum S is the output of an OR gate; the inputs are obtained from AND gates; inversion of A and B is necessary. Equation 15-1 also indicates that the carry C is simply the output of an AND gate. The corresponding logic circuit is shown in Fig. 15.3c.

There may be several different Boolean expressions for any given logic statement and some will lead to better circuit realizations than others. Algebraic manipulation of Eq. 15-1 yields

$$\begin{aligned} \bar{A}B + A\bar{B} &= \overline{(A + \bar{B})(\bar{A} + B)} && \text{(DeMorgan's theorems)} \\ &= \overline{A\bar{A} + AB + \bar{A}\bar{B} + B\bar{B}} && \text{(Multiplication)} \\ &= \overline{\bar{A}\bar{B} + AB} && (A\bar{A} = 0 \text{ and } B\bar{B} = 0) \\ &= (A + B)(\bar{A} + \bar{B}) && \text{(DeMorgan's theorems)} \\ &= (A + B)\overline{AB} && \text{(DeMorgan's theorems)} \end{aligned}$$

Looking again at the truth table (Fig. 15.3b), we see that another interpretation is: "S is 1 if (A OR B) is 1 AND (A AND B) is NOT 1." Therefore binary addition can be expressed as:

$$S = (A + B)\overline{AB} \quad \text{and} \quad C = AB \qquad (15\text{-}2)$$

The synthesis of this circuit, working backward from the output, is shown in Fig. 15.4. This circuit is better than that of Fig. 15.3c in that fewer logic elements are used and the longest path from input to output passes through fewer *levels*. The ability to optimize logic statements is an essential skill for the logic designer.

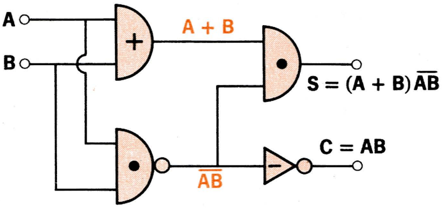


Figure 15.4 Another half-adder circuit.

The Exclusive–OR *Gate*

The function $(A + B)\overline{AB}$ in Equation 15-2 is called the **Exclusive–OR** operation. As indicated by the truth table, it can be expressed as: "**A OR B** but **NOT A AND B**." The alternate form (Eq. 15-1a) $\bar{A}B + A\bar{B}$ is called an "inequality comparator" because it provides an output of 1 if A and B are not equal.

Example 8

Show that the inverse of the inequality comparator is an "equality comparator" and synthesize a suitable circuit.

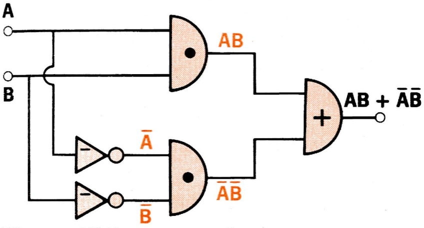


Figure 15.5 An equality comparator.

Algebraic manipulation of the inverse function using DeMorgan's theorems yields

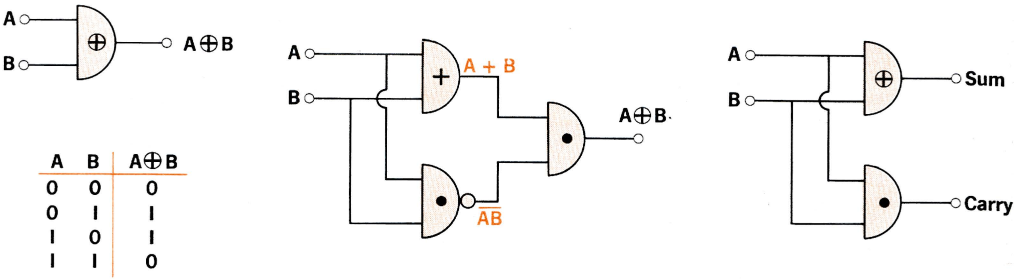
$$\overline{\bar{A}B + A\bar{B}} = (A + \bar{B})(\bar{A} + B) = AB + \bar{A}\bar{B} \qquad (15\text{-}3)$$

This is an equality comparator in that the output is 1 if A and B are equal. This function is highly useful in digital computer operation. Straight-forward synthesis results in the circuit of Fig. 15.5.

The **Exclusive–OR** gate is used so frequently that it is represented by the special symbol ⊕ defined by

$$A \oplus B = (A + B)\overline{AB} \qquad (15\text{-}4)$$

One realization of the **Exclusive–OR** gate is shown in Fig. 15.6b. Assuming that this gate is available as a logic element, the half-adder takes on the simple form of Fig. 15.6c. As a further simplification we can treat the half-adder as a discrete logic element and represent it by a rectangular block labeled **HA**.



A	B	A⊕B
0	0	0
0	1	1
1	0	1
1	1	0

(a) Symbol and truth table (b) Circuit (c) Use in half-adder

Figure 15.6 An Exclusive-OR circuit and its application.

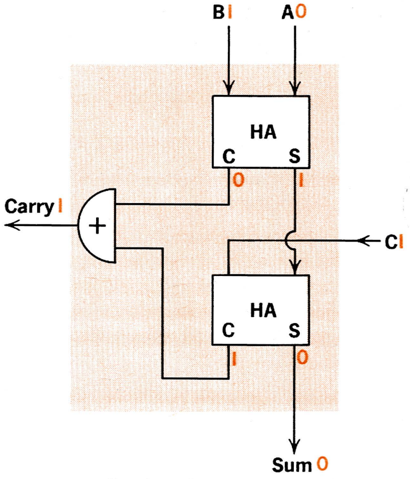
The Full-Adder

In adding two binary digits or *bits*, the half-adder performs the most elementary part of what may be highly sophisticated computation. To perform a complete addition, we need a *full-adder* capable of handling the carry input as well. The addition process is illustrated in Fig. 15.7a, where C is the carry from the preceding column. Each carry of 1 must be added to the two digits in the next column, so we need a logic circuit capable of combining three inputs. This operation can be realized using two half-adders and an OR gate. The values

```
 A     1011
+B   + 1110
      11001
```

```
C      1110
A     01011
B     01110
Sum   11001
Carry 01110
```

(a) Addition process



(b) Circuit realization

Figure 15.7 Design of a full-adder.

shown in Fig. 15.7b correspond to the third column of the addition. The operation of the full-adder can be seen more clearly by construction of a truth table (Exercise 15).

To perform the addition of "4-bit words," i.e., numbers consisting of four binary digits, we need a half-adder for the first column and a full-adder for each additional column. In the *parallel binary adder,* the input to the half-adder consists of digits A_1 and B_1 from the first column. The input to the next unit, a full-adder, consists of carry C_1 from the first unit and digits A_2 and B_2 from the second column. The sum of the two numbers is represented by C_4 S_4 S_3 S_2 S_1. In subtraction, **NOT** gates provide the 2's complement of the subtrahend (see footnote, p. 390) and the process is reduced to addition.

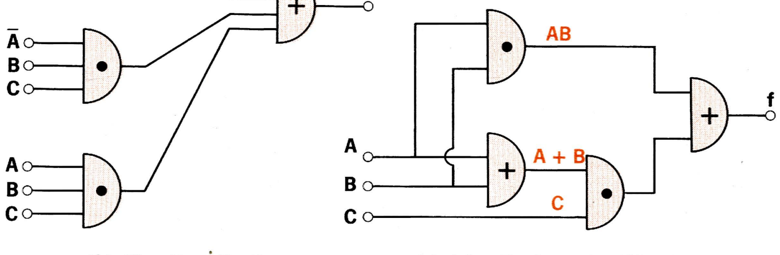
A Design Procedure

A common problem in logic circuit design is to create a combination of gates to realize a desired function. The basic approach is to proceed from a statement of the function to a truth table and then to a Boolean expression of the function. The experienced logic designer then manipulates the Boolean expression into the simplest form. The realization of the final expression in terms of **AND**, **OR**, and **NOT** gates is straightforward.

For increased reliability on a spacecraft, triple sensing systems are used; no action is taken unless at least two of the three systems call for action. The truth table of the required *vote taker* is shown in Fig. 15.8a. Since the function is YES

A	B	C	f
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

(a) Truth table



(b) Circuit realization

(c) After Boolean simplification

Figure 15.8 Design of a vote taker.

(1) only when a majority of the inputs are YES, the Boolean expression must contain a product term for each row of the truth table in which the function is 1. Since the function is YES for any one *or* more of these rows, the Boolean expression is

$$f = AB\bar{C} + A\bar{B}C + \bar{A}BC + ABC \tag{15-5}$$

Assuming that the complement of each variable is available, as is true in most computers, the straightforward realization is a combination of four **AND** gates feeding an **OR** gate (Fig. 15.8b).

If the complements were not available, eight logic elements would be required and simplification of the circuit would be desirable. First, the Boolean function of Equation 15-5 can be expanded into

$$f = AB\bar{C} + A\bar{B}C + \bar{A}BC + ABC + ABC$$

since $ABC + ABC = ABC$. Factoring by the distribution rule yields

$$f = AB(\bar{C} + C) + C(AB + A\bar{B} + \bar{A}B)$$

Since $\bar{C} + C = 1$ and $AB + A\bar{B} + \bar{A}B = A + B$, the function becomes

$$f = AB + C(A + B)$$

This function requires only four logic elements (Fig. 15.8c).

NAND *Gate Synthesis*

In practice, logic systems are designed to employ a single type of gate so that electrical characteristics are nearly identical throughout. Since the transistor is basically an inverter, DTL and TTL units are inherently **NAND** gates. In the case where all logic functions are to be synthesized using **NAND** gates only, a straightforward procedure is as follows:

1. Use the truth table to write a function in the *sum-of-products* form

   $$f = P_1 + P_2 + \cdots + P_n$$

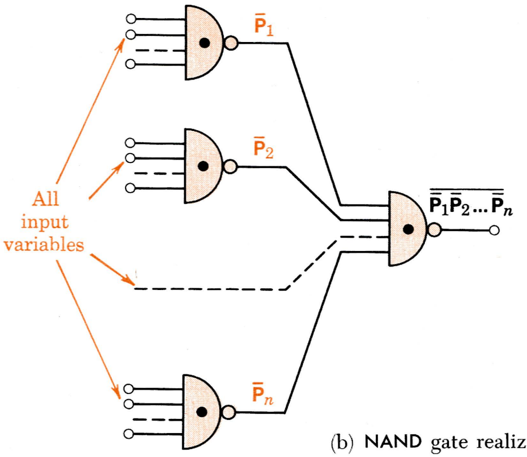
   where each P is the product of the input variables in a row in which the output is 1.
2. In the product corresponding to a given row, invert (place a bar over) each variable for which the truth table value is 0.
3. Use DeMorgan's theorem to convert the function to

   $$f = \overline{\bar{P}_1\bar{P}_2 \cdots \bar{P}_n}$$

4. Synthesize the circuit using **NAND** gates as indicated in Fig. 15.9 and demonstrated in Example 9.

A	B	C	D	f	
0	0	0	0	0	
0	0	0	1	1	P_1
·	·	·	·	·	
0	1	0	1	1	P_2
·	·	·	·	·	
1	1	1	1	1	P_n

(a) Function truth table


$\bar{P}_1$
$\bar{P}_2$
All input variables
$\overline{\bar{P}_1\bar{P}_2\ldots\bar{P}_n}$
$\bar{P}_n$

(b) NAND gate realization

Figure 15.9 NAND gate synthesis.

Example 9

From the truth table of the full-adder, extract the rows in which Sum = 1 and draw a partial truth table. Design the sum circuit of a full-adder using only NAND gates.

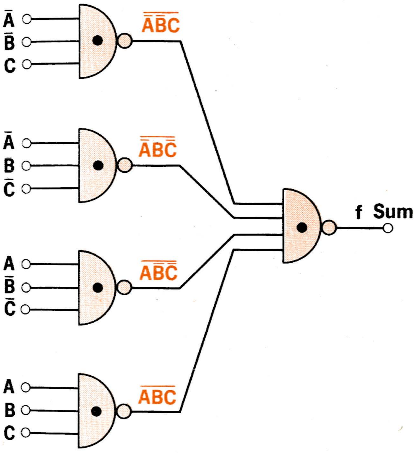

$\bar{A}$
$\bar{B}$
C
$\overline{\bar{A}\bar{B}C}$
$\bar{A}$
B
$\bar{C}$
$\overline{\bar{A}B\bar{C}}$
A
$\bar{B}$
$\bar{C}$
$\overline{A\bar{B}\bar{C}}$
A
B
C
$\overline{ABC}$
f Sum

Figure 15.10 Full-adder using NAND gates.

A	B	C	Sum
0	0	1	1
0	1	0	1
1	0	0	1
1	1	1	1

Following the general procedure for NAND synthesis:

1. $ABC + ABC + ABC + ABC$
2. $f = \bar{A}\bar{B}C + \bar{A}B\bar{C} + A\bar{B}\bar{C} + ABC$
3. $f = \overline{(\overline{\bar{A}\bar{B}C})(\overline{\bar{A}B\bar{C}})(\overline{A\bar{B}\bar{C}})(\overline{ABC})}$
4. Assuming that the complement of each input is available (as is true in most computers), the realization is as shown in Fig. 15.10. If the complements are not available, three NAND gates can be used for the NOT operations.

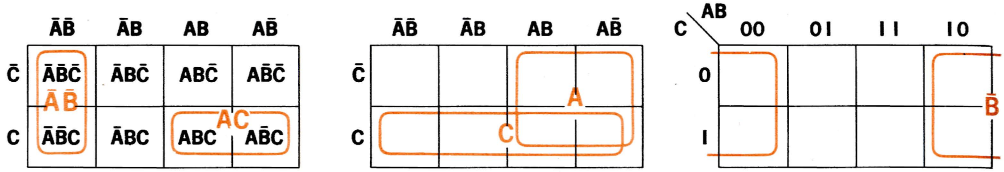
The carry of a full-adder is identical with the vote taker (Fig. 15.12b). An IC full-adder would include complement circuits, the sum circuit, and the carry circuit.

Minimization by Mapping

In creating a logic circuit to realize a desired function, the designer seeks the optimum form. The criterion may be maximum speed (fewest logic levels) or minimum cost (fewest gate leads since number of leads determines the cost of manufacture and the cost of assembly). We have seen how Boolean algebra can be used in deriving simpler logic expressions. If the truth table is available or if the logic function is expressed as a "sum of products," the designer can go directly to the minimal expression by the mapping technique suggested by Maurice Karnaugh.

Karnaugh Maps

The *map* of the general logic function of three variables is shown in Fig. 15.11a.



(a) 2–square terms (b) 4–square terms (c) Another 4–square term

Figure 15.11 Mapping a three-variable logic function.

Each square in the map corresponds to one of the eight possible combinations of the three variables. The order of the columns is such that *combinations in adjacent squares differ only in the value of one variable.* Therefore, 2-square groups are independent of one variable, i.e., $f_1 = \bar{A}\bar{B}\bar{C} + \bar{A}\bar{B}C = \bar{A}\bar{B}$ and $f_2 = ABC + A\bar{B}C = AC$. These relations are easy to derive using Boolean algebra, but they are *obvious by inspection* of the Karnaugh map.

The concept can be extended to groupings of four adjacent squares as shown in Fig. 15.11b, where the labels are omitted from the squares. The 4-square cluster outlined in color is independent of both B and C and the 4-square in-line group is independent of both A and B, i.e., $f_3 = \bar{A}\bar{B}C + \bar{A}BC + ABC + A\bar{B}C = C$. Enlarging groups by overlapping simplifies the terms. Note that the map is "continuous" in that the last column on the right is "adjacent" to the first column on the left. The 4-square group in Fig. 15.11c is just equal to $\bar{B}$.

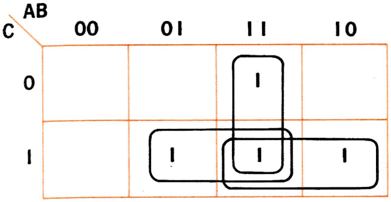
The standard labeling scheme for Karnaugh maps (Fig. 15.11c) is convenient for mapping from a truth table. Each square in the map corresponds to a row in the truth table. A specific logic function is *mapped* by placing a 1 in each square for which the function is 1. When this has been done, possible simplifications are easily recognized. (See Example 10.)

Mapping in Four Variables

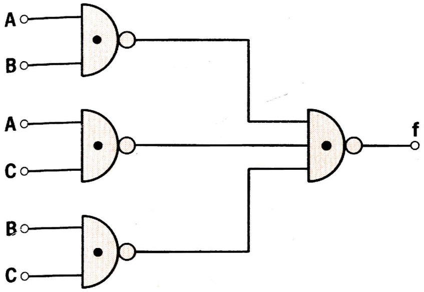
The Karnaugh technique is even more valuable in simplifying functions in four variables. (For more than four variables, other techniques are usually more

Example 10

Map the vote-taker function and simplify the circuit realization, if possible.


AB
C
00
01
11
10
0
1
1
1
1
1

(a) Map of f



(b) Minimal realization

Figure 15.12 Simplification by Karnaugh mapping.

From the truth table of Fig. 15.8a, 1s are placed in the squares corresponding to rows in the truth table for which the function is 1, representing

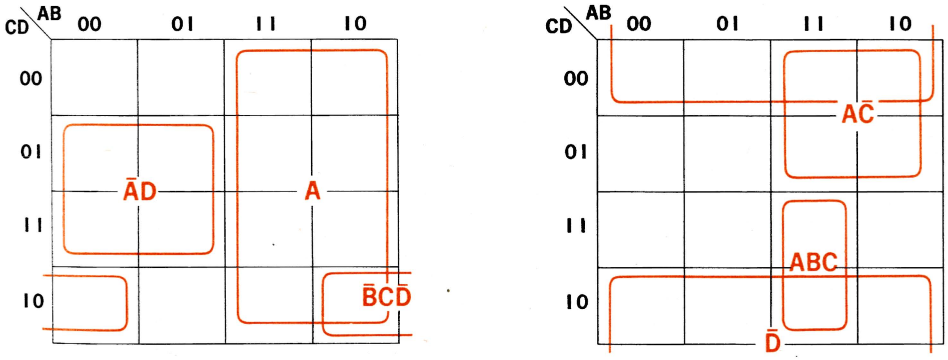
$$f = \bar{A}BC + A\bar{B}C + AB\bar{C} + ABC$$

All the 1s can be included in three overlapping 2-square groups; ∴ the complete function can be represented by

$$f = AB + AC + BC$$

Note that in the circuit realization of Fig. 15.12b the number of gate leads has been reduced to 9 input + 4 output = 13 from the 21 in Fig. 15.8b. This vote taker could be realized in a single package.

convenient.) As indicated in Fig. 15.13, 2-square groups are independent of one variable, 4-square groups are independent of two variables, and 8-square groups are independent of 3 variables. Note that the standard labeling scheme provides


CD
AB
00
01
11
10
00
01
11
10
$\bar{A}D$
A
$\bar{B}C\bar{D}$
CD
AB
00
01
11
10
00
01
11
10
$A\bar{C}$
ABC
$\bar{D}$

(a) (b)

Figure 15.13 Examples of grouping on four-variable maps.

that adjacent rows differ by only one complement bar and the bottom row is *adjacent* to the top row. The four corner squares form a combination that is a little difficult to visualize.

There are detailed rules for finding the minimal expression from a Karnaugh map,† but some general guidelines will suffice for our purposes:

Include all 1s in groups of eight, four, two, or one.
Groups may overlap; larger groups result in simpler terms.
Of the possible combinations of terms, select the simplest.

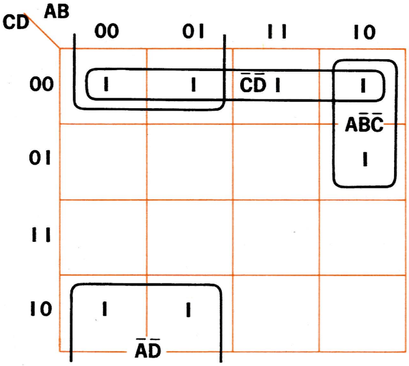
The technique is illustrated in Example 11.

Example 11

Map the function

$$f = \overline{AB}(\overline{C + D}) + \overline{A}C\overline{D} + AB\overline{C}\overline{D} + A\overline{B}\overline{C}D$$

and obtain a minimal sum-of-products expression.



AB
CD
00
01
11
10
00
01
11
10
1
1
$\overline{C}\overline{D}$ 1
1
$A\overline{B}\overline{C}$
1
1
1
$\overline{A}\overline{D}$

Figure 15.14 Four-variable simplification.

As an alternative to forming the truth table, let us map the function by considering the factors individually. $\overline{AB}$ limits the first term to the 00, 01, and 10 columns and $(\overline{C + D})$ corresponds to the first row, implying three 1s as shown in Fig. 15.14. The $\overline{A}C\overline{D}$ term is independent of B implying a 2-square group in the lower left-hand corner. The four-variable terms imply 1s in the 1100 and 1001 squares.

All the 1s can be included in the two 4-square and one 2-square groups encircled.

$$\therefore f = \overline{A}\overline{D} + \overline{C}\overline{D} + A\overline{B}\overline{C}$$

Other expressions are possible, but none will include fewer, simpler terms.

Two additional comments should be made. In some circuits certain combinations of inputs never occur; such *don't care* combinations may be mapped as Xs and considered as either 0s or 1s, whichever provides the greatest simplification. In other circuits, the simplest realization results from implementing $\overline{f}$ as a sum of products and then inverting to obtain f. (See Exercise 25.)

IC Logic Gates

In the full-adder of Fig. 15.10, five NAND gates are required just to add two binary digits. A pocket calculator providing 10-place accuracy must handle decimal numbers up to 9,999,999,999 or $\sim 2^{34}$. Therefore, 34 bits are required for the binary addition, that is, 33 full-adders and a half-adder for a total of about 175 **NAND** gates. Since the arithmetic unit must also provide for multiplication (repeated addition), subtraction (adding the 2s complement), and division (repeated subtraction) the number of identical gates required in even a small calculator is very large. The great advantages of the integrated circuit—low cost, small size, and identical performance from gate to gate—are apparent.

---

† See pp. 69 ff of the Peatman reference at the end of Ch. 14.

Even in a simple digital control system, the basic gates are used in large numbers. Instead of being fabricated individually, several gates are constructed in a single chip and packaged as a unit. The number of gates in a package is usually limited by the number of pin connections available. In the flat-pack or the dual-in-line package, there are 14 pins; if one is ground and one is power supply, there are 12 available for input and output. As examples, standard items include: single 8-input **NAND/NOR**, triple 3-input **NAND/NOR**, dual 4-input **OR/NOR**, and quad 2-input **AND/OR/INVERT**.

At the next level of integration, there are MSI (medium scale integration) units containing from a dozen to a hundred individual gates per chip. Available as standard items are flip-flops of various design, 2- and 4-bit full-adders, full-subtractors, shift registers, and decade counters. Manufacturer's principal specifications include fan-out, propagation delay, and power dissipation. The circuit designer must select units that match his other components and perform the desired functions.

Computer Organization

An electronic digital computer is a vast assemblage of simple logic gates and memory devices organized to perform complex calculations at high speed by automatically processing information in the form of electrical pulses. The information consists of *data* and *instructions;* a complete set of instructions is called a *program.*

The functional organization of a computer is illustrated in Fig. 15.15. Although

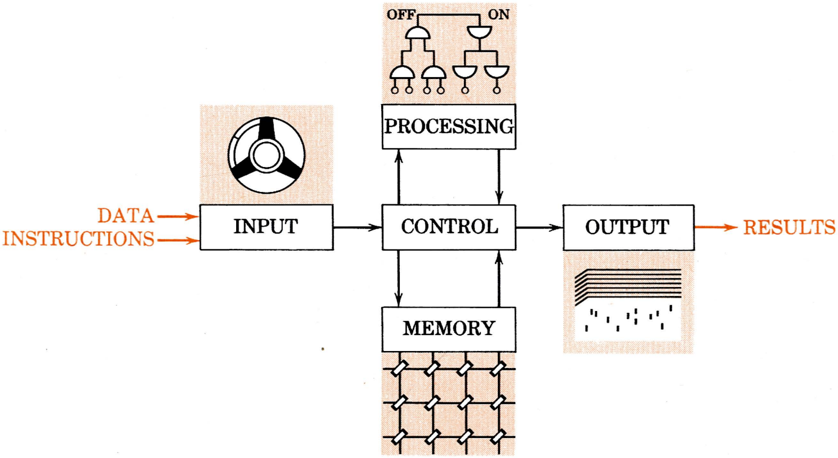


Figure 15.15 Functional components of a digital computer.

the designs of individual computer components vary widely, all general-purpose computers follow the same general plan.

Input

Instructions and data are entered into the machine through an input device that reads data on punched cards, paper tape, or magnetic tape and converts it to electrical waveforms. All communication from the operator to the machine is through the input device. Most commonly this is a magnetic tape recorder (see Fig. 27.24).

Memory

Each instruction and datum is "stored" in the computer memory until it is "retrieved" for use. Flip-flops are used for temporary storage of data and instructions in active registers. The working computer memory usually consists of an array of magnetic cores, although solidstate memories are becoming more common.

Each tiny doughnut-shaped core may be magnetized in either of two directions, corresponding to logical 0 and 1, by the momentary passage of current through a pair of wires threading the core. The unique pair of wires threading each core identifies the *address* of that core. The state of a core is read by sending a CLEAR pulse through the wires and sensing whether a change in magnetization occurs. A typical *core plane* contains 128 × 128 = 16,384 cores; a *stack* of 60 such planes provides a storage capacity of nearly a million bits of information.

Control

The flow of instructions, data, and results is governed by the control unit, which sets the various logic gates, feeds the numerical data, and provides the clock pulses that regulate the speed of all operations. Instructions are received from the input device and stored in one part of the memory; data are stored in another part. In accordance with the program, the control unit provides for a desired

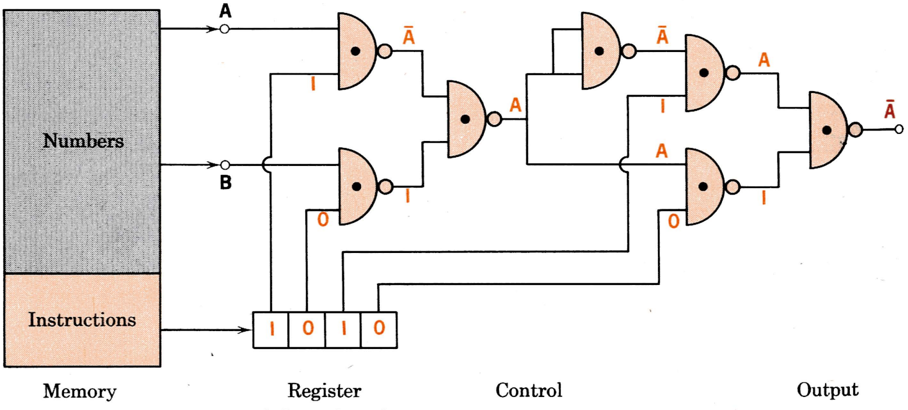


Figure 15.16 Control of data flow by instructions in a register.

sequence of arithmetic operations and retrieves the numerical data from memory.

The operation of the control unit is illustrated in the elementary example of Fig. 15.16. Here stored numbers **A** and **B** are to be operated on in accordance with instructions in the form of a four-bit word. The instruction is drawn from memory and placed in a register associated with a logic circuit of seven **NAND** gates. The instruction **1010**, as shown, sets the logic gates so that **A** is selected and **NOT A** is passed on to the next unit or returned to memory for further processing. Other instructions set the logic gates for different operations (see Exercise 27).

Processing

The basic operation in computation is the addition of two binary numbers. The processing or *arithmetic* unit consists of logic gates and function circuits arranged to perform addition and the associated registers for temporary storage of data or subtotals. In adding a column, the first number is taken from memory and placed in a register, the second number is added, a third is then added to the sum of the first two, and so on. In *parallel* addition, there is a separate adder for each bit and all orders are added in one operation. In *serial* addition, numbers are added one bit at a time by a single full-adder; the numbers and the sum are stored in shift registers (Fig. 14.22).

Output

The computer presents its results to the operator in printed, punched, or graphical form by means of an output device. Usually the limiting factor in data processing is the speed of the output device. Modern line printers can convert electrical signals to typed characters at the rate of 20 lines per second. Electrostatic printing permits even higher speeds. Punching cards or paper tape is relatively slow, but the results are readily stored for reuse. Plotters are available for graphical presentation of calculations where this is desirable.

SUMMARY

- Binary data values **0** and **1** correspond to OFF and ON switch positions.
  Conversions from binary to decimal and vice versa follow simple rules.
  Binary arithmetic follows the same rules as decimal arithmetic.

- Boolean algebra is used to describe and to manipulate binary relations.
  Boolean theorems are summarized in Tables 15-1, -2, and -3.
  DeMorgan's theorems are particularly helpful in **NAND** synthesis.

- To analyze a logic circuit, trace through the circuit from input to output, noting the function obtained at each step.
  To synthesize a logic circuit from logic statement or truth table, use Boolean algebra or Karnaugh mapping to simplify the function, and realize it with standard elements.
  To synthesize a circuit in **NAND** gates, a general procedure is available.

- Useful computer components and their *logic* functions include:

Exclusive-OR gate: $f = (A + B)\overline{AB}$ (or $f = A\overline{B} + \overline{A}B$)
Half-adder: $S = (A + B)\overline{AB}$ $C = AB$
Equality comparator: $f = AB + \overline{A}\,\overline{B}$

Basic gates and functions are available at low cost in IC form.

- A digital computer is a complex array of logic and memory elements organized to perform binary calculations at high speed.
Computers include input, memory, control, processing, and output sections.

REVIEW QUESTIONS

1. Why is the binary system useful in electronic data processing?
2. How can decimal numbers be converted to binary? Binary to decimal?
3. What is Boolean algebra and why is it useful in digital computers?
4. Which of the Boolean theorems in 0 and 1 contradict ordinary algebra?
5. Explain the association and distribution rules for Boolean algebra.
6. State DeMorgan's theorems in words. When are they useful?
7. How can a logic circuit be analyzed to obtain a logic statement?
8. How can a logic statement be synthesized to create a logic circuit?
9. What function is performed by a half-adder? An Exclusive-OR gate?
10. Explain how binary numbers are added in a parallel adder.
11. Explain how we can design a NAND circuit from a truth table.
12. Explain the principle behind the numbering of columns in a Karnaugh map.

EXERCISES

1. (a) Write in binary the following decimals: 3, 10, 20, 31.
   (b) Write in decimal the following binaries: 00001, 00100, 01010, 11010.
2. (a) Write in binary the following decimals: 7, 13, 21, 39, 75.
   (b) Write in decimal the following binaries: 00101, 01100, 11110, 101011.
   (c) Convert decimals 0.875 and 0.65 to binaries.
3. Given $A = \frac{65}{4}$ and $B = \frac{3}{2}$, convert A and B to decimal and binary form and evaluate $A \times B$ and $A \div B$ in each form. Check your answers.
4. Repeat Exercise 3 for $A = \frac{27}{8}$ and $B = \frac{18}{4}$.
5. Use the method outlined in the footnote on p. 390 to perform 36 − 14.
6. Repeat Exercise 5 to perform 11001 − 10110.
7. Use truth tables to prove the following Boolean theorems:
   (a) $A + 1 = 1$,
   (b) $AB + AC = A(B + C)$,
   (c) $A + B = B + A$, (d) $A(BC) = (AB)C$.
8. Use truth tables to prove the following Boolean theorems:
   (a) $A + (B + C) = (A + B) + C$,
   (b) $(A + B)(A + C) = A + BC$,
   (c) $A(B + C) = AB + AC$.
9. Use Boolean theorems to prove the following identities:
   (a) $ABC + AB\overline{C} = AB$,
   (b) $(A + B)(\overline{A} + C) = AC + \overline{A}B$.
   (c) $A(\overline{A} + B) = AB$,
   (d) $AB + \overline{A}C = (A + C)(\overline{A} + B)$,
   (e) $(A + C)(A + D)(B + C)(B + D) = AB + CD$.
10. The function $f = A \cdot B$ is to be realized

using only **NOR** gates. Use DeMorgan's theorems to express f in terms of $\overline{C + D}$ where C and D can be expressed in terms of A and B. Draw the necessary logic circuit and check by constructing the truth table.

11. Analyze the logic circuit of Fig. 15.17 and construct the truth table. Simplify the logic statement and check your result against the truth table.

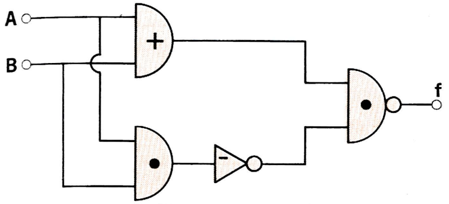


Figure 15.17

12. Repeat Exercise 11 for the logic circuit of Fig. 15.18.

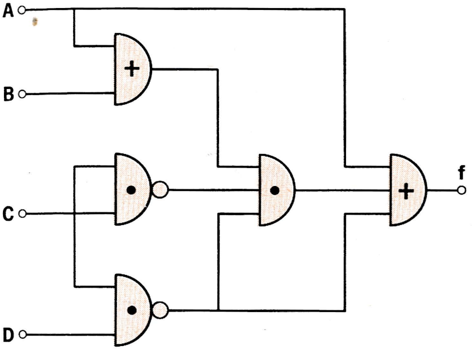


Figure 15.18

13. Devise a "half-subtractor."
  (a) Construct the truth table providing **A**, **B**, **Borrow**, and **Difference** columns.
  (b) Synthesize the circuit using only the five basic gates.
  (c) Synthesize the circuit assuming **Exclusive–OR** gates are available.

14. To provide a comparison of the most common logic operations, construct a table with inputs **A** and **B** and outputs **AND**, **NAND**, **OR**, **NOR**, **Exclusive–OR**, and **Equality Comparator**.

15. Construct a truth table for the full-adder of Fig. 15.7.

16. Devise a parallel binary adder to add the binary numbers $A_4A_3A_2A_1$ and $B_4B_3B_2B_1$. Draw a labeled block diagram using full-adder units. Label all inputs and outputs to illustrate the binary addition of 3 and 9; repeat for the addition of 8 and 17.

17. Synthesize logic circuits to realize the following functions as written:
  (a) $\overline{(A + B) + \overline{AB}} \cdot \overline{AB}$.
  (b) $\overline{A + B} \cdot \overline{AB} + AB$.

18. Synthesize the following logic functions using **NAND** gates only:
  (a) **AND, OR, NOR.**
  (b) **Exclusive–OR.**
  (c) **Equality Comparator.**

19. Synthesize the following logic functions using **NOR** gates only:
  (a) **AND, OR, NAND.**
  (b) **Exclusive–OR.**

20. Given the logic function

$$f = \overline{AB + \bar{A}\bar{B}} + \bar{A}B$$

Assuming the complements are available:
  (a) Simplify the function using DeMorgan's theorem and synthesize it using the basic gates.
  (b) Construct the truth table and use **NAND** circuit synthesis to realize the function.

21. Map the following functions and find the minimal sum-of-products form:
  (a) $ABC + \bar{A}\bar{B}\bar{C}$
  (b) $AB + AC$
  (c) $(A + \bar{C})(\bar{B} + \bar{C})$
  (d) $ABC + BC + A\bar{B}C$

22. Evaluate the combination of the four corner squares of a four-variable Karnaugh map.

23. Map the following functions and find the minimal sum-of-products form:
  (a) $ABC\bar{D} + A\bar{B}C + \bar{B}\bar{C}$
  (b) $AB + \bar{A}\bar{B}CD + A\bar{B}C$
  (c) $AB(CD + \bar{C}\bar{D})$
  (d) $\bar{A}(C + D) + A\bar{B}C + A\bar{B}\bar{C}\bar{D}$

24. Map the function $f = AB(C + D)$, simplify it to the minimal sum-of-products form, and synthesize it using **NAND** gates only.

25. The Karnaugh map of Fig. 15.19 contains three "don't care" combinations.

CD \ AB	00	01	11	10
00	×	1	1	×
01			1	1
11	1	1		1
10	1	1		×

Figure 15.19

(a) Determine a minimal sum-of-products expression for f.

(b) Draw a new map for $\overline{f}$ and obtain a minimal expression.

(c) Assuming complements are available, compare the number of gates, gate leads, and logic levels for the two realizations.

26. Design (i.e., sketch the construction of) an integrated circuit realization of the DTL NAND gate in Fig. 14.7b following the form of Fig. 13.24.

27. Refer to Fig. 15.16 and interpret the instruction:

(a) 1001

(b) 0110

(c) 0101

PROBLEMS

1. An electric light is to be controlled by three switches. The light is to be ON whenever switches A and B are in the same position; when A and B are in different positions, the light is to be controlled by switch C.

(a) Draw up a truth table for this situation.

(b) Represent the light function f in terms of A, B, and C.

(c) Simplify the function and design a practical switching circuit.

2. One way of checking for errors in data processing is to "code" each word by attaching an extra bit so that the number of 1's in each "coded" word is *even*, i.e., the words all have *even parity*. Assuming that the probability of *two* errors in one word is *negligible*, the coded word can be checked for even parity at various points in the process. Assuming Exclusive-OR gates are available, design a *parity checker* for coded 4-bit words such that the output is 0 for even parity.

3. In the HP 35 pocket calculator, each character of the 15-digit display consists of 7 light-emitting diodes arranged as shown in Fig. 15.20. For example, decimal "1" is displayed when segments *c* and *d* are lit.

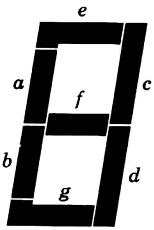


Figure 15.20 Seven-segment decimal display.

(a) Determine how each decimal from 0 to 9 could be represented and draw the corresponding segment display.

(b) Draw up the truth table showing the state of each segment for the decimal numbers.

(c) If the elementary binary counter of Fig. 14.20 is used to drive the display, segment *b* should be lit for certain combinations of outputs from flip-flops A through D. Write the logic expression governing segment *b*.

4. The "comparator," widely used in computers, has an output of 0 unless the inputs are identical. Synthesize, using NAND gates only, a circuit that "compares" the two-digit binary numbers A_1A_2 and B_1B_2.

Chapter 16

Amplifier Bias and Large-Signal Performance

Practical Amplifiers
Biasing Circuits
Power Amplifiers
Other Amplifiers

The two major functions performed by electronic devices are switching and amplifying. We have studied the use of diodes and transistors in switching and other nonlinear applications. Using the characteristics of two-port devices described in Chapters 10 and 13 and the circuit theory of Part I, we are now ready to study the analysis and design of various types of electronic amplifiers.

The design of electronic amplifiers to meet critical performance, weight, and cost specifications requires a great deal of knowledge and judgment. Electrical engineers will gain the detailed knowledge in subsequent courses and the judgment from practical experience. Other engineers and scientists will be more concerned with the use of existing amplifiers; if they do any design (in connection with instrumentation, for example), it will be under circumstances where optimum performance is not essential and a simplified approach is satisfactory.

Our purpose here is to discuss the factors that influence amplifier performance, illustrate the basic analysis techniques, and present methods for designing simple amplifiers and predicting their performance. Our approach is first to look at the practical considerations in amplifier operation and devise circuits for maintaining the proper operating conditions, then to analyze the performance of one type of power amplifier, and finally to describe briefly some other important types of amplifiers.

Practical Amplifiers

Amplifier Classification

There are many ways of describing amplifiers. A *single-stage* amplifier consists of one amplifying element and the associated circuitry; in general, several such elements are combined in a *multistage* amplifier. In a sound reproduction system, for example, the first stages are *small-signal* voltage (or current) amplifiers designed to amplify the output of a phonograph pickup, a few millivolts, up to a signal of several volts. The final stage is a *large-signal* or *power* amplifier that supplies sufficient power, several watts, to drive the loudspeaker.

Such amplifiers are called *audio* amplifiers if they amplify signals from, say, 30 to 15,000 Hz. In measuring structural vibrations, temperature variations, or the electrical currents generated within the human body, very low-frequency signals are encountered; to handle signals from zero frequency to a few cycles per second, *direct-coupled* amplifiers are used. For comparison, the *video* amplifier in a television receiver must amplify picture signals with components from 30 to 4,000,000 Hz.

A video amplifier is a *wide-band* amplifier that amplifies equally all frequencies over a broad frequency range. In contrast, a *radiofrequency* amplifier for the FM broadcast band (around 100 MHz) is *tuned* to select and amplify the signal from one station and to reject all others. In this chapter we are primarily interested in simple untuned audio power amplifiers.

Amplification and Distortion

The terms amplification and gain are used almost interchangeably. For sinusoidal signals or for a particular sinusoidal component of a periodic signal the voltage gain is

$$\mathbf{A}_V = \frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = A\,e^{j\theta} \tag{16-1}$$

where $\mathbf{A}_V$ is the complex ratio of two phasors. In a *linear amplifier* A and θ are independent of signal amplitude and frequency, and the output signal is a replica of the input signal.†

If there is *distortion* in the amplifier, the output is not a replica of the input. In Fig. 16.1, output is not proportional to input and there is *nonlinear* or *amplitude* distortion. In other words, A is not a simple constant. As a result of amplitude distortion, there are frequency components in the output that are not present in the input. Fourier analysis of the output would reveal the presence of a "second harmonic," a component of twice the frequency of the "fundamental" input signal.‡ Amplitude distortion usually occurs when excessively large signals are applied to nonlinear elements such as transistors.

At the other extreme is distortion due to *noise*, random signals unrelated to the input. If the input signal is too small, the output consists primarily of noise and is not a replica of the input. The "snow" that appears on a television screen when only a weak signal is available is a visual representation of noise. One source of noise is the random thermal motion of electrons in the amplifier circuit elements. The *shot effect* of individual electrons arriving at the plate of a vacuum tube or recombining at a junction is another source of noise. Noise is of greatest importance in input stages where signal levels are small; any noise introduced there is amplified by all subsequent stages. The *dynamic range* of any amplifier is bounded at one end by the level at which signals are obscured by noise and at the other by the level at which amplitude distortion becomes excessive.

---

† If $\theta = k\omega$, a time lag is introduced, but there is no phase distortion.

‡ The calculation of harmonic distortion is outlined in Chapter 18 of J. Millman and C. Halkias: *Integrated Electronics*, McGraw-Hill Book Co., New York, 1972.

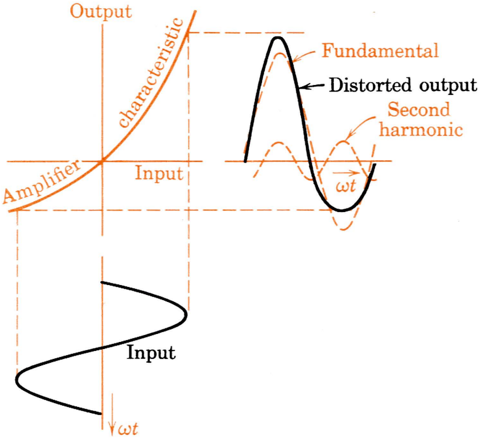


Figure 16.1 Amplitude distortion.

In Fig. 16.2, the *frequency response curve* of an audio amplifier indicates that there is *frequency distortion;* all frequencies (within a finite band) are not amplified equally. In other words, gain A in Eq. 16-1 is a function of frequency. A signal consisting of a fundamental at 1 kHz, a tenth harmonic at 10 kHz, and a hundredth harmonic at 100 kHz would have a different waveform after amplification. No amplifier is completely free from frequency distortion.

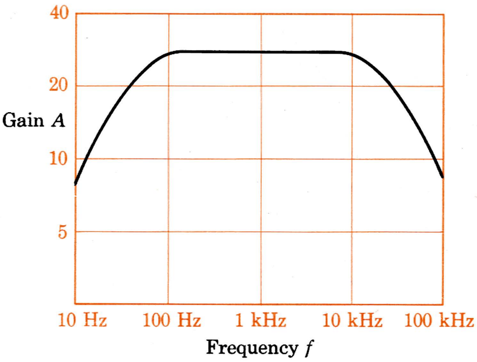
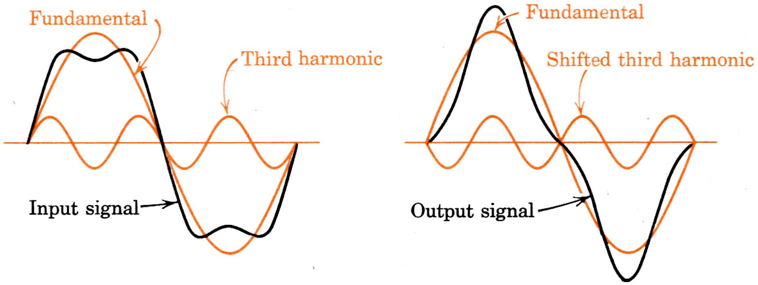


Figure 16.2 Frequency distortion.

If θ is a function of frequency, the relative amplitudes of the signal components may be unchanged but the relative phase positions are shifted. As shown in Fig. 16.3, such *phase distortion* changes the shape of the output wave. The eye is sensitive to such distortion but the ear is not; a human ordinarily cannot distinguish between the two signals. On the other hand, the ear is quite sensitive to amplitude or frequency distortion.

Phase and frequency distortion are caused by circuit elements such as capacitive and inductive reactances that are frequency dependent. Some transistor parameters also are frequency dependent. In the design of untuned or wide-band amplifiers special steps are taken to reduce the variation in gain with frequency.



(a) Input signal (b) Output signal

Figure 16.3 Phase distortion.

Practical Considerations

Some of the practical considerations in amplifier design are illustrated in the simplified two-stage audio amplifier of Fig. 16.4.

Biasing. Transistors and FETs are maintained at the proper operating point by dc power supplies and biasing networks. In portable units, batteries are used; in other units, the readily available 60-Hz current is rectified and filtered to provide the necessary direct current. In Fig. 16.4, the battery V_{CC} supplies the collector voltages to transistors TR1 and TR2 and, through resistors R_B and R_D, the base biasing currents. When several different bias voltages are required, they are obtained from a voltage divider or by series *dropping* resistors.

Coupling. The voltage generated in the phonograph pickup is coupled to TR1 by a combination of R_{VC} and C_C. Variable resistor R_{VC} provides volume control by voltage-divider action. Transistor TR1 is coupled to the second stage by means of collector resistor R_C and coupling capacitor C_C. Such an *RC* coupling circuit develops a useful signal output across R_C and transfers it to the input of the next stage, but dc voltages and currents are blocked by C_C.

The loudspeaker is coupled to TR2 by transformer T. Transformer coupling is more expensive than *RC* coupling, but it is more efficient and it permits impedance matching for improved power transfer. The combination of R_D and C_D is a *decoupling filter* to prevent feedback of amplified signals to the low-level input stage. Such feedback is likely to occur when the common battery supplying several stages begins to age and develops an appreciable internal resistance.

Load Impedance. If possible, the load impedance of an untuned amplifier is made purely resistive to minimize the variation in gain with frequency. In tuned amplifiers, the load impedance is usually a parallel resonant circuit. For voltage amplifiers, large values of R_C are desirable; but large R_C requires large supply voltages, so the selected value is a compromise.

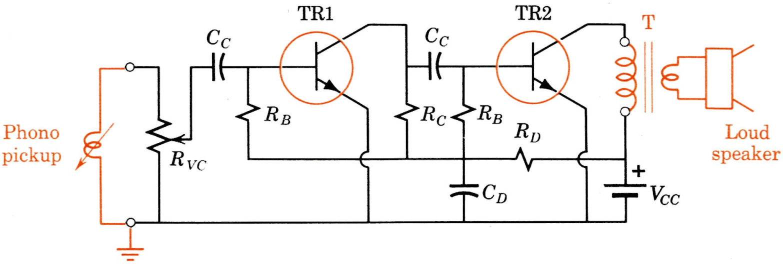


Figure 16.4 A simplified two-stage phonograph amplifier.

Input and Output Impedance. When several amplifier stages are connected in *cascade,* the output characteristics of one stage are influenced by the input characteristics of the next. For ac signals, R_C, the load resistance of TR1, is effectively shunted by the biasing resistance R_B and also by the input resistance of TR2. Ideally the input impedance of an amplifier stage should be high to minimize "loading" of the preceding stage, and the output impedance should be low for efficient power transfer.

Unintentional Elements. Figure 16.4 is a wiring diagram and shows components, not circuit elements. Even a short, straight conductor can store a small amount of energy in the form of a magnetic field or an electric field (between the conductor and ground or a metal chassis). The magnetic field effect is important only at extremely high frequencies; the electric field effect can be represented by a *wiring capacitance.*

In the same way we represent the energy storage due to a difference in potential between the gate and source of a MOSFET by an *equivalent capacitance* (Fig. 17.9). At a transistor junction there is a charge separation and a potential difference across the transition region (see Fig. 11.10); a change in junction potential causes a change in charge distribution ($q = Cv$) and this effect is represented by a *junction capacitance* C_j.

Transistors exhibit another charge storage effect. Under equilibrium conditions, the base must be electrically neutral. Some of the charge from majority carriers (holes in a *pnp* transistor) diffusing from the emitter into the base is neutralized by carriers (electrons in *pnp*) from the base connection. A change in emitter-base potential causes a transfer of charge into or out of the base; this effect is represented by a *diffusion capacitance* C_d. A circuit model used for analyzing amplifier performance at high frequencies must include these unintentional elements and, therefore, it will be quite different from the wiring diagram of Fig. 16.4.

Biasing Circuits

The objective in biasing is to *establish* the proper operating point and to *maintain* it despite variations in temperature and variations among individual devices of the same type. Furthermore, this objective is to be achieved without adversely affecting the desired performance of the circuit. The biasing problem is difficult because of:

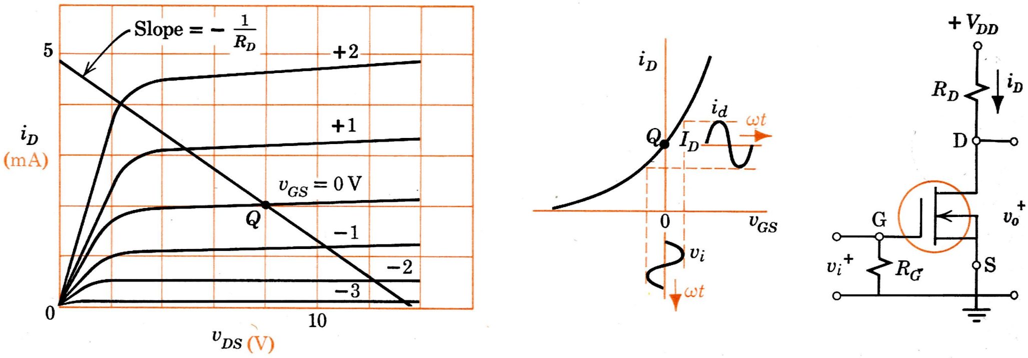
(a) The wide variations in device parameters expected among mass-produced transistors,
(b) The complicated interrelations among transistor variables, and
(c) The inherent sensitivity of semiconductor devices to temperature.

Because of its importance, much attention has been given to this problem and as a result many ingenious circuits are available for the designer's use.

FET Biasing

Field-effect transistors are used primarily as "small-signal" amplifiers because the parabolic transfer characteristic introduces amplitude distortion if the signals are large. (See Example 2 in Chapter 13.) If the excursions along the load line are small (Fig. 13.17b), the biasing requirements are not critical.

DE MOSFET. Biasing a depletion-or-enhancement (DE) MOSFET is extremely simple because of its unique transfer characteristic that allows a quiescent operating point at $V_{GS} = 0$. The transfer characteristic of a DE MOSFET (Fig. 13.5b) is repeated in Fig. 16.5b. To obtain this characteristic, the drain-source channel must be biased by a dc voltage that places v_{DS} in the normal operating region between turn-on and breakdown where drain current i_D is nearly independent of v_{DS}. The necessary circuit (Fig. 16.5c) contains a voltage supply V_{DD}, a drain resistor R_D across which signal voltages can be developed, and a gate resistor R_G that allows any charge that might build up on the highly insulated gate to "leak" off.

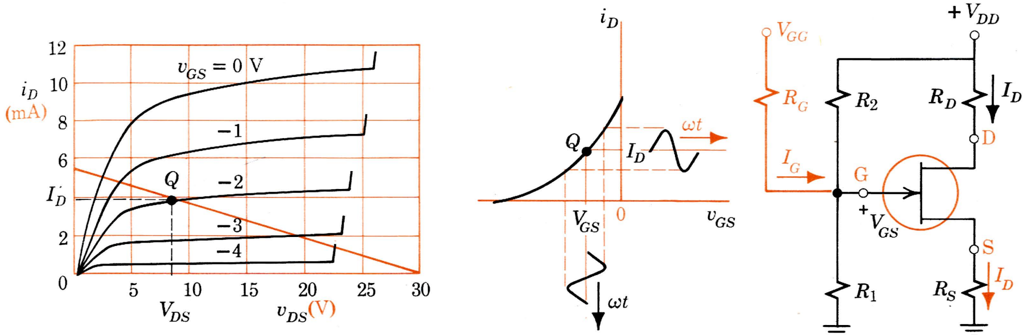


(a) *i-v* characteristics (b) Transfer characteristic (c) Circuit diagram

Figure 16.5 A DE MOSFET amplifier.

As indicated by the transfer characteristic, a small signal voltage applied to the input produces a signal current at the output. The output voltage $v_o = V_{DD} - i_D R_D$ contains a signal component that is an amplified replica of the input voltage.

JFET. Biasing a depletion-mode JFET is complicated by the fact that positive values of v_{GS} are not permitted; ac signals must be superimposed on a steady negative dc value V_{GS}.† (For nomenclature, refer again to Table 13-1, p. 348.) A possible Q point is shown in Fig. 16.6a and b. A further complication is that V_{GS} for a given I_D varies from sample to sample and with temperature.



(a) i–v characteristics (b) Transfer characteristic (c) Bias circuit

Figure 16.6 Biasing a depletion-mode JFET.

To provide the necessary negative value of V_{GS}, one possibility would be to use a negative source V_{GG}, but this would require a second battery. Another simple possibility is to place a resistor R_S between source and ground so that the source is at a potential $+I_D R_S$ with respect to ground; if then the gate is at ground potential, $V_{GS} = -I_D R_S$ as desired. A more complicated but better biasing scheme is shown in Fig. 16.6c, where R_1 and R_2 form a voltage divider to place the gate at the desired potential with respect to ground. As a first step in the analysis, replace V_{DD}, R_1, and R_2 by the Thévenin equivalents

$$V_{GG} = \frac{R_1}{R_1 + R_2} V_{DD} \quad \text{and} \quad R_G = \frac{R_1 R_2}{R_1 + R_2} \tag{16-2}$$

Summing voltages around the gate "loop" yields

$$\Sigma V = 0 = V_{GG} - I_G R_G - V_{GS} - I_D R_S \tag{16-3}$$

Disregarding the negligible gate current through a reverse-biased junction, the voltage equation reduces to

---

† Vacuum tubes resemble JFETs insofar as bias requirements and circuits are concerned. See pp. 404 ff of Reference 1 at the end of this chapter.

$$I_D R_S = V_{GG} - V_{GS} \tag{16-4}$$

For a specified Q point (I_D, V_{GS}) and chosen values of V_{GG} and R_G, the required values of R_S, R_1, and R_2 are easily calculated from

$$R_S = \frac{V_{GG} - V_{GS}}{I_D} \qquad R_2 = R_G \frac{V_{DD}}{V_{GG}} \qquad R_1 = \frac{R_G R_2}{R_2 - R_G} \tag{16-5}$$

Furthermore, by making V_{GG} large compared to $|V_{GS}|$, the effect of any shift in V_{GS} is reduced.

Example 1

The JFET of Fig. 16.6 is to be operated at a quiescent point defined by $I_D = 4$ mA, $V_{DS} = 8$ V, and $V_{GS} = -2$ V. Design an appropriate biasing circuit with $V_{DD} = 30$ V.

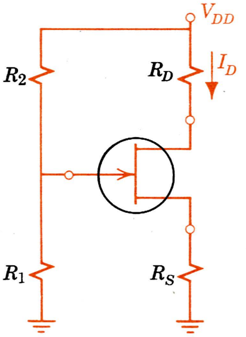


Figure 16.7 JFET bias design.

Using the voltage-divider circuit of Fig. 16.7, assume $V_{GG} = 12$ V so that $V_{GG} - V_{GS}$ is large compared to V_{GS} for stability, and assume $R_G = 6$ MΩ to keep the input resistance high. Then, by Eqs. 16-5,

$$R_S = \frac{V_{GG} - V_{GS}}{I_D} = \frac{12 - (-2)}{4 \text{ mA}} = 3.5 \text{ k}\Omega$$

$$R_2 = R_G \frac{V_{DD}}{V_{GG}} = 6\frac{30}{12} = 15 \text{ M}\Omega$$

$$R_1 = \frac{R_G R_2}{R_2 - R_G} = \frac{6 \times 15}{15 - 6} = 10 \text{ M}\Omega$$

Summing voltages around the drain "loop" yields

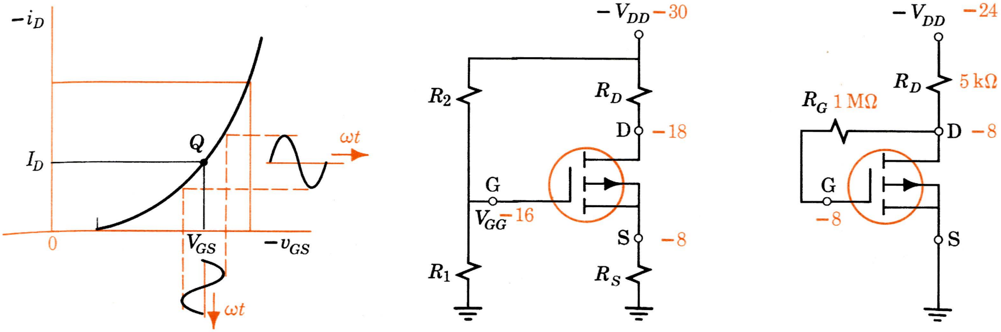
$$\Sigma V = 0 = V_{DD} - I_D R_D - V_{DS} - I_D R_S$$

For $V_{DS} = 8$ V,

$$\begin{aligned} R_D &= (V_{DD} - V_{DS} - I_D R_S)/I_D \\ &= (30 - 8 - 14)/4 \text{ mA} = 2 \text{ k}\Omega \end{aligned}$$

Note that a 20% shift in V_{GS} would cause only a 3% change in $V_{GG} - V_{GS}$ and in I_D.

Enhancement MOSFET. For an enhancement-only MOSFET (Fig. 13.3), normal operation requires a V_{GS} of the proper polarity to attract the holes or electrons necessary for conduction in the channel. If the voltage-divider bias circuit of Fig. 16.8b is used, V_{GG} is specified to provide the proper magnitude



(a) Transfer characteristic (b) Voltage-divider bias (c) Drain–feedback bias

Figure 16.8 Biasing a p-channel enhancement-only MOSFET.

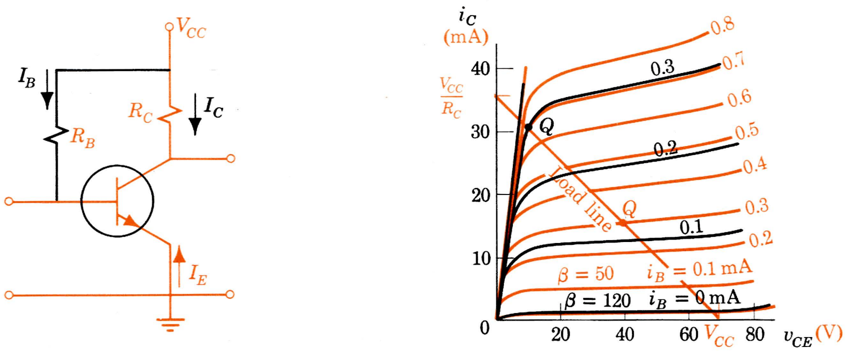
and polarity of $V_{GS} = V_{GG} - V_S = V_{GG} - I_D R_S$. (Some representative voltages are shown in color.)

The simpler bias circuit shown in Fig. 16.8c is satisfactory in many situations. Since the gate current is negligible, $V_{GS} = V_{DS}$ and operation is in the normal region for $|V_{GS}| > |V_T|$. (The design of such a bias circuit was illustrated in Example 1 of Chapter 13; the design values are shown in color on Fig. 16.8c.) In this *drain-feedback bias* circuit, any change in drain current due to changes in device or circuit parameters is *fed back* to the gate circuit in such a way as to compensate for the parameter change. For example, if I_D increases for any reason, V_{DS} decreases and the corresponding decrease in V_{GS} tends to decrease I_D. Feedback resistor R_G must be very large to minimize the degrading effect of signal current flowing from the output back to the input.

BJT Biasing

In normal operation of a bipolar transistor the emitter-base junction is forward biased and the collector-base junction is reverse biased. Since in the common-emitter configuration these bias voltages have the same polarity, a single battery can supply both through an appropriate circuit. Here again, however, the variation in parameters among mass-produced transistors and the inherent sensitivity of semiconductors to temperature complicate the biasing problem, particularly in the large-signal amplifiers to be discussed in the next section. After demonstrating that the simplest solution is unsatisfactory, we analyze one commonly used bias circuit and outline an approximate design procedure.

Fixed-Current Bias. One possibility is to provide the desired dc base current from V_{CC} as in Fig. 16.9. For a forward-biased emitter the junction voltage is



(a) Circuit diagram (b) Collector characteristics

Figure 16.9 Fixed-current bias and effect of β change.

a fraction of a volt and may be neglected in comparison to V_{CC}. On this basis, the quiescent base current is $I_B \cong V_{CC}/R_B$ and the operating point would be determined if the collector characteristics were known precisely.

But the characteristics are not known precisely. For steady or dc values (see Table 13-1) Eq. 13.6 becomes

$$I_C = \frac{\alpha}{1-\alpha} I_B + \frac{I_{CBO}}{1-\alpha} = \beta I_B + (1+\beta) I_{CBO} \tag{16-6}$$

Even if I_B is fixed, collector current I_C may vary widely with the large variations in β and I_{CBO} expected when transistors are interchanged. As shown in Fig. 16.9b, a value of I_B that places the quiescent point in the linear region of a transistor with $\beta = 50$, pushes operation into the saturation region of another transistor of the same type with $\beta = 120$. Also, I_C will vary widely with temperature changes because β increases nearly linearly with temperature and I_{CBO} increases exponentially with temperature.

Thermal Runaway. The power loss in a transistor is primarily at the collector junction because the voltage there is high compared to the low voltage at the forward-biased emitter junction. If the collector current I_C increases, the power developed tends to raise the junction temperature. This causes an increase in I_{CBO} and β and a further increase in I_C which tends to raise the temperature higher. In a transistor operating at high temperature (because of high ambient temperature or high developed power) a regenerative heating cycle may occur which will result in *thermal runaway* and possibly destruction of the transistor.

For equilibrium, the power developed in the transistor is equal to the power (heat) dissipated to the surroundings. The ability to dissipate heat by conduction is proportional to the difference between junction and ambient temperatures and inversely proportional to the thermal resistance of the conducting path. For a constant ambient temperature, the ability to dissipate heat increases as the junction temperature goes up. Equilibrium will be upset and thermal runaway may occur when the increase in power to be dissipated is greater than the increase in dissipation ability due to the increase in junction temperature. One way to avoid this cumulative effect is to cool the collector junction; power transistors utilize the heat conducting capacity of the metal chassis on which they are mounted or special heat-radiating fins. Another way is to use an effective biasing circuit.

Self Bias. The ingenious self-biasing circuit of Fig. 16.10 decreases the effect of changes in β or temperature on the quiescent operating point. Its operation is based on the fact that the critical variable to be stablized is the collector current rather than the base current. The combination of R_1 and R_2 constitutes a voltage divider (as in Fig. 16.6c) to bring the base to the proper potential to forward bias the emitter junction. If I_C tends to increase, perhaps because of an increase in β due to a rise in temperature, the current $I_C + I_B$ in R_E increases, raising the potential of the emitter with respect to ground. This, in turn, reduces V_{BE}, the forward bias on the base-emitter junction, reduces the base current and, therefore, limits the increase in I_C. In other words, any increase in collector current is fed back to the base circuit and modifies the bias in such a way as to oppose a further increase in I_C.

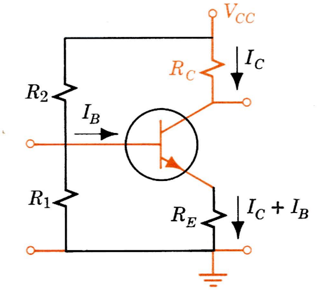
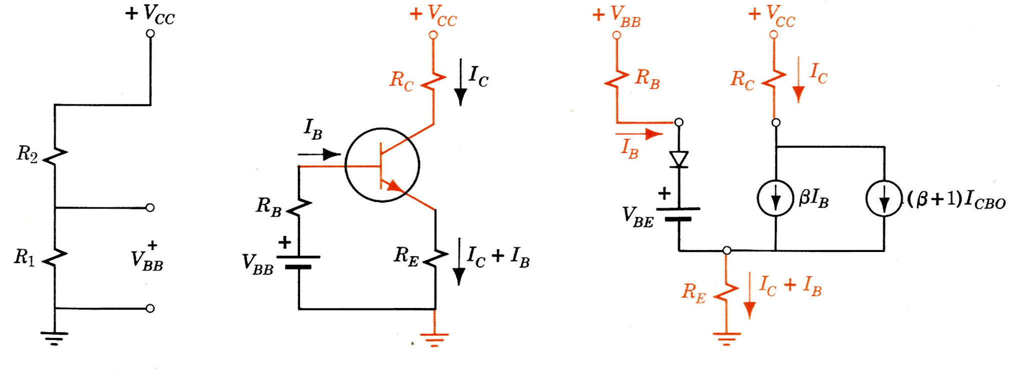


Figure 16.10 A self-biasing circuit.

Quantitative analysis of the circuit is simplified if the voltage divider (Fig.



(a) Voltage divider (b) Thevenin equivalent (c) Circuit model

Figure 16.11 Analysis of the self-biasing circuit.

16.11a) is replaced by its Thévenin equivalent where

$$V_{BB} = \frac{R_1}{R_1 + R_2} V_{CC} \quad \text{and} \quad R_B = \frac{R_1 R_2}{R_1 + R_2} \tag{16-7}$$

After replacing the transistor by a dc model (Fig. 16.11c), Kirchhoff's voltage law around the base loop yields

$$V_{BB} - I_B R_B - V_{BE} - (I_C + I_B) R_E = 0 \tag{16-8}$$

From Eq. 16-6, represented in the model by two current sources† in parallel,

$$I_B = \frac{I_C}{\beta} - \frac{\beta + 1}{\beta} I_{CBO} \tag{16-9}$$

Substituting in Eq. 16-8 and solving,

$$I_C = \frac{V_{BB} - V_{BE} + \left(\frac{\beta + 1}{\beta}\right) I_{CBO}(R_B + R_E)}{R_E + \frac{R_B + R_E}{\beta}} \tag{16-10}$$

This is a general equation for collector current in a self-biasing circuit where V_{BE}, β, and I_{CBO} are device parameters that are known imprecisely and that are sensitive to temperature change. Typically, $|V_{BE}|$ decreases 2.5 mV/°C, β may vary by 6 : 1 and increases linearly with temperature, I_{CBO} may vary by 10 : 1 and approximately doubles for every 10°C increase above 25°C. (See Example 2.)

Example 2 demonstrates that a properly designed self-biasing circuit is effective in stabilizing I_C despite variations in device parameters. Equation 16-10 reveals the design criteria; to stabilize operation, each term containing a variable parameter should be made "small" with respect to a circuit constant. Since the terms are to be "small," rough approximations are acceptable in evaluating the terms; for example, for $\beta \geq 20$, $(\beta + 1)/\beta \cong 1$, and for $R_B \geq 5R_E$, $R_B + R_E \cong R_B$. With these approximations in mind, the design criteria are:

To make I_C independent of I_{CBO}, make

$$\left(\frac{1 + \beta}{\beta}\right) I_{CBO}(R_B + R_E) \cong I_{CBO} R_B \ll V_{BB} - V_{BE} \tag{16-11}$$

To make I_C independent of β, make

$$\frac{R_B + R_E}{\beta} \cong \frac{R_B}{\beta} \ll R_E \tag{16-12}$$

---

† The current generator βI_B is called a "controlled current source" since the output is determined by a variable in another part of the circuit.

Example 2

The β of individual specimens of a silicon transistor varies from 30 to 180. If V_{BE} may vary from 0.5 to 0.9 V and I_{CBO} may vary from 1 to 10 nA, predict the extreme variation in I_C in a self-biasing circuit where $R_1 = 10\ \text{k}\Omega$, $R_2 = 100\ \text{k}\Omega$, $R_E = 1\ \text{k}\Omega$, $R_C = 5\ \text{k}\Omega$, and $V_{CC} = 22$ V.

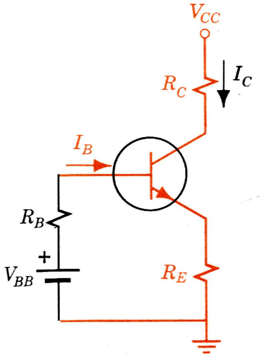


Figure 16.12 Temperature limitation.

By Eqs. 16-7, the Thévenin equivalents are

$$R_B = \frac{R_1 R_2}{R_1 + R_2} = \frac{10 \times 100}{10 + 100} = 9.1\ \text{k}\Omega$$

$$V_{BB} = \frac{V_{CC} R_1}{R_1 + R_2} = \frac{22 \times 10}{10 + 100} = 2\ \text{V}$$

By Eq. 16-10, the collector current (Fig. 16.12) is

$$I_C = \frac{V_{BB} - V_{BE} + \left(\dfrac{\beta + 1}{\beta}\right) I_{CBO}(R_B + R_E)}{R_E + \dfrac{R_B + R_E}{\beta}}$$

For $\beta = 30$, $V_{BE} = 0.9$ V, and $I_{CBO} = 1$ nA (the "worst case"),

$$I_C = \frac{2 - 0.9 + \left(\dfrac{31}{30}\right) 10^{-9}(10.1 \times 10^3)}{1 \times 10^3 + (10.1 \times 10^3)/30}$$

$$= \frac{1.1 + 0.00001}{(1 + 0.34) \times 10^3} = 0.75\ \text{mA}$$

For $\beta = 180$, $V_{BE} = 0.5$ V, and $I_{CBO} = 10$ nA,

$$I_C = \frac{2 - 0.5 + \left(\dfrac{181}{180}\right) 10^{-8}(10.1 \times 10^3)}{1 \times 10^3 + (10.1 \times 10^3)/180}$$

$$= \frac{1.5 + 0.0001}{(1 + 0.056) \times 10^3} = 1.42\ \text{mA}$$

For these extreme variations, I_C shifts only 0.7 mA. Note that for a typical silicon transistor, $(1 + \beta)I_{CBO} = I_{CEO}$ is negligible in bias calculations.

To make I_C independent of V_{BE}, make

$$V_{BE} \ll V_{BB} \tag{16-13}$$

The effect of temperature on the Q point can be predicted in terms of changes in V_{BE} and I_{CBO}. Assuming the circuit is designed to be independent of β (Eq. 16-12), the collector current is approximated by

$$I_C = \frac{V_{BB} - V_{BE} + I_{CBO} R_B}{R_E} \tag{16-14}$$

Assuming V_{BB}, R_B, and R_E are constant (actually ±10% variations are to be expected), the *change* in I_C will be

$$\Delta I_C = \frac{-\Delta V_{BE} + \Delta I_{CBO} R_B}{R_E} \tag{16-15}$$

The relative sensitivity of silicon and germanium is evaluated in Example 3.

Example 3

A silicon transistor with $V_{BE} = 0.7$ V and $I_{CBO} = 10$ nA and a germanium transistor with $V_{BE} = 0.3$ V and $I_{CBO} = 5\ \mu$A (typical room temperature values) are used in the circuit of Example 2. Predict the effect on collector current of a 50° increase in operating temperature.

The change in V_{BE} will be

$$\Delta V_{BE} = (-2.5\ \text{mV/°C})50\text{°C} \cong -0.1\ \text{V}$$

I_{CBO} will change by the factor

$$2^{\Delta T/10} = 2^{50/10} = 2^5 = 32$$

and the new values will be

$$\text{Si}: I_{CBO} = 32 \times 10 = 320\ \text{nA}$$

$$\text{Ge}: I_{CBO} = 32 \times 5 = 160\ \mu\text{A}$$

By Eq. 16-15, the *change* in I_C will be

$$\text{Si}: \Delta I_C = \frac{+0.1 + 310 \times 10^{-9} \times 9.1 \times 10^3}{10^3}$$

$$= \frac{0.1 + 0.003}{10^3} \cong 0.1\ \text{mA}$$

$$\text{Ge}: \Delta I_C = \frac{+0.1 + 155 \times 10^{-6} \times 9.1 \times 10^3}{10^3}$$

$$= \frac{0.1 + 1.4}{10^3} = 1.5\ \text{mA}$$

Bias Design, Approximate Method

As demonstrated in Example 3, silicon transistors are relatively unaffected by temperature changes and the small change in collector current is primarily due to ΔV_{BE}. Germanium transistors are sensitive to temperature changes and the principal factor is ΔI_{CBO}.

In stabilizing against variations in β, Eq. 16-12 indicates that R_B should be small compared to βR_E. In practice, the ratio $\beta R_E/R_B$ is limited. R_B must not be too small because it appears directly across the input and tends to divert part of the signal current. Also, R_E must not be too large because part of the dc supply voltage V_{CC} appears across R_E (Fig. 16.13). As R_E is increased, less voltage is available for developing an output signal across R_C, and the collector circuit efficiency is reduced. (To prevent an undesirable ac voltage variation across R_E, it is customarily by-passed by capacitor C_E that offers low impedance at all signal frequencies. Coupling capacitors C_C isolate the dc bias currents from the signal source and the following stage.)

The designer must select values of R_1, R_2, and R_E to provide optimum performance; but the criteria are different in every situation and skill and experience are essential to an optimum design. If large temperature changes are not expected and if less than optimum performance is acceptable, the following procedure† is satisfactory:

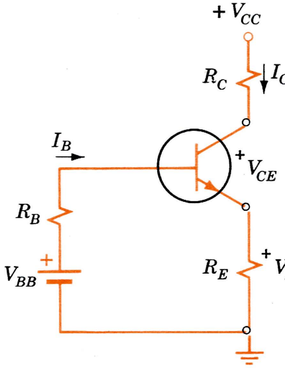


Figure 16.13 Bias design.

1. Select an appropriate nominal operating point (I_C, I_B, and V_{CE}) from the manufacturer's data (see Fig. 16.9b).
2. Arbitrarily assume that $V_E = I_E R_E \cong I_C R_E \cong 3$ V, say, and solve for R_E.
3. Select V_{CC} and R_C.
   (a) If V_{CC} is specified, $R_C \cong (V_{CC} - V_{CE} - 3)/I_C$ (16-16)
   (b) If R_C is specified, $V_{CC} \cong 3 + V_{CE} + I_C R_C$ (16-17)
4. Arbitrarily select R_B equal to $\beta_{\min} R_E/10$. (16-18)
   (For $\beta = 50$, a 10% change in β will cause a 1% change in I_C.)
5. Calculate $V_{BB} = I_B R_B + V_{BE} + (I_B + I_C)R_E$ (16-19)
   (Lacking other information, use $I_B = I_C/\beta$ and assume $V_{BE} \cong 0.7$ V for silicon or 0.3 V for germanium)
6. Calculate R_2 and R_1 (see Eqs. 16-7) from

$$R_2 = R_B \frac{V_{CC}}{V_{BB}} \quad \text{and} \quad R_1 = \frac{R_2 R_B}{R_2 - R_B} \tag{16-20}$$

The exact sequence in which these steps are taken can be modified to fit a given situation. (See Example 4, p. 427.)

Power Amplifiers

To drive a loudspeaker or a recording instrument requires an appreciable amount of power. In analyzing or designing a *power amplifier,* large signals and the accompanying nonlinearity must be considered. Transistors (or tubes) capable of delivering high power output are expensive and therefore careful attention to efficient amplifier design is justified.

---

† C. L. Searle, et al., *Elementary Circuit Properties of Transistors,* John Wiley & Sons, New York, 1964, p. 167.

The basic design problem is to provide the desired power output with a stable circuit that uses the amplifying device efficiently and safely. Let us first consider the permissible operating region, see how operating point and load resistance are selected, derive expressions for power and efficiency, and then analyze a common type of audiofrequency amplifier that uses a power transistor.

Permissible Operating Region

An electronic amplifier must operate without introducing excessive distortion and without exceeding the voltage, current, and power limitations of the device. The bipolar junction transistor provides a large linear region and is commonly used where large signal excursions are required. The permissible operating region can be indicated on the output characteristics of the transistor as in Fig. 16.14.

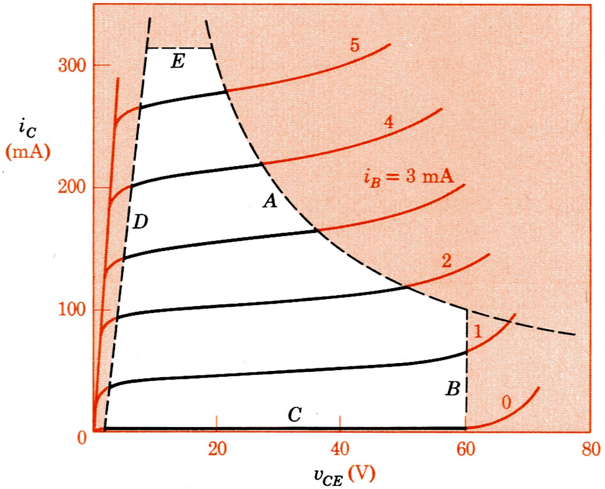


Figure 16.14 Permissible operating region of a transistor.

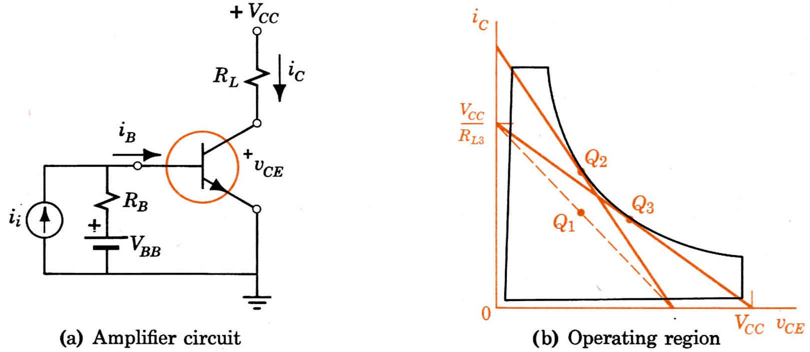
Line A represents the maximum allowable power dissipation P_D for the transistor; this is an hyperbola defined by $V_{CE}I_C = P_C = P_D$, where P_D is established by the manufacturer. Operation above this line may damage the device. Line B reflects the fact that at high collector voltages the avalanche effect causes a rapid increase in collector current and the curves become nonlinear.

For any electronic device there are regions of excessive nonlinearity that should be avoided. Line C bounds the region in which collector current is approaching zero and a further decrease in signal value (i_B) does not produce a corresponding decrease in output current. Line D bounds the saturation region in which a further increase in signal does not produce a corresponding increase in output current. Line E indicates an arbitrary limit within which the transistor manufacturer guarantees his specifications rather than a maximum allowable current.

Operating Point and Load Line

The purpose of the biasing arrangement is to locate and maintain operation in the permissible region. Within this limitation we wish to obtain maximum power output. In general, the load line should lie below the maximum dissipation curve, and its slope (determined by the load resistance) should reflect a compromise between large signals and low distortion.

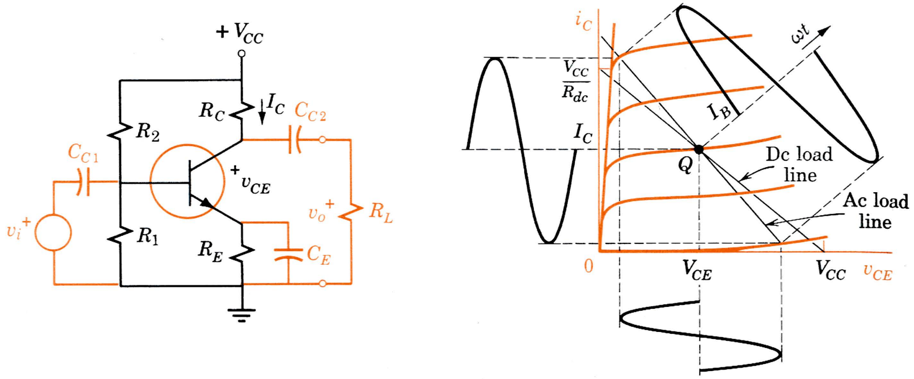
The factors influencing the choice of Q point and load resistance are shown in Fig. 16.15, where R_L is assumed to be in the collector circuit. Quiescent point Q_1 is well below the maximum dissipation curve and does not permit maximum voltage and current swings. A second possibility is Q_2 where the load line for $R_{L2} < R_{L1}$ is drawn tangent to the hyperbola; this permits the same voltage swing as Q_1 and a larger current swing. The load line for $R_{L3} > R_{L1}$, tangent to the hyperbola at Q_3, permits the same current swing as Q_1 and a larger voltage swing. Operation at any point along the maximum dissipation curve between Q_2 and Q_3 will permit approximately the same signal power output. The choice will depend on practical factors such as the available supply voltage or the resistance of the load.



(a) Amplifier circuit (b) Operating region

Figure 16.15 Location of the load line for a simple amplifier.

The practical amplifier of Fig. 16.16a includes some additional complexities. Since the capacitances are open circuits for dc bias currents, the operating point and *dc load line* are determined from the equivalent dc circuit where $R_{dc} = R_C + R_E$. For maximum voltage swings, operation should be near the midpoint of the load line or $V_{CC} \cong 2V_{CE}$. Since $V_{CE} = V_{CC} - I_C R_{dc}$,



(a) Amplifier circuit (b) Load line construction

Figure 16.16 Operation of a large-signal amplifier.

$$R_{dc} = R_C + R_E \cong \frac{V_{CE}}{I_C} \tag{16-21}$$

a typical design condition.

The ac signals "see" a different circuit. R_B is chosen large enough so that most of the signal current flows into the base. C_E is an ac short circuit; therefore R_E is effectively removed. V_{CC} is a steady dc source and there can be no ac voltage across it; therefore the upper end of R_C is effectively grounded. C_{C2} is an ac short circuit, therefore R_C and R_L are effectively in parallel across the output voltage v_o, which is the signal component of v_{CE}. The ac load line defined by

$$R_{ac} = R_C \| R_L = \frac{R_C R_L}{R_C + R_L} \tag{16-22}$$

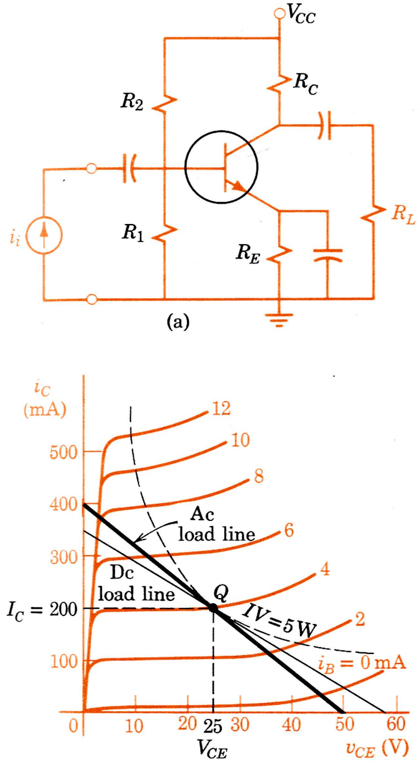
is shown in Fig. 16.16b. For this circuit $R_{dc} > R_{ac}$ and the optimum location of the Q point is at the midpoint of the ac load line.

Transformer Coupling

As illustrated in Example 4, the supply battery voltage V_{CC} must exceed the average collector voltage V_{CE} by the average voltage drop across the collector resistor $I_C R_C$. Furthermore, most of the signal power developed is dissipated in

Example 4

The audiofrequency amplifier of Fig. 16.17 employs the silicon power transistor for which characteristics are shown. The manufacturer specifies $I_C(\text{max}) = 1$ A, $V_{CE}(\text{max}) = 50$ V, $P_D(\text{max}) = 5$ W (case at 25°C), and $\beta_{\text{min}} = 30$. Design the amplifier for maximum power to $R_L = 750\ \Omega$; an approximate design is satisfactory. (Assume capacitances are "large.")



(a)

(b)

Figure 16.17 Amplifier design.

To select the operating point, the maximum dissipation line $I_C V_{CE} = 5$ W is sketched in as shown. The ac load line should be tangent to the hyperbola and there is a range of slopes with nearly equal output powers. The high-voltage region is judged to be slightly more linear than the high-current region, so the decision is made to use the full allowable voltage and let the current be limited by dissipation. At $V_{CE} = 50/2 = 25$ V and $I_C = 0.2$ A,

$$P_C = V_{CE} I_C = 25 \times 0.2 = 5 \text{ W}$$

and the corresponding ac load resistance is

$$R_{\text{ac}} = \frac{V_{CE}}{I_C} = \frac{25}{0.2} = 125\ \Omega$$

$$\therefore R_C = \frac{1}{\dfrac{1}{R_{\text{ac}}} - \dfrac{1}{R_L}} = \frac{1}{0.008 - 0.0013} = 150\ \Omega$$

To design the bias circuit, we assume a voltage $V_E = 3$ V; the emitter resistance is

$$R_E \cong \frac{V_E}{I_C} = \frac{3}{0.2} = 15\ \Omega$$

Then the necessary supply voltage is

$$V_{CC} = 3 + V_{CE} + I_C R_C = 3 + 25 + 30 = 58 \text{ V}$$

To minimize the effect of changes in β, let

$$R_B = \frac{\beta_{\text{min}} R_E}{10} = \frac{30 \times 15}{10} = 45\ \Omega$$

Next, we read $I_B = 4$ mA, assume $V_{BE} = 0.7$ V for silicon, and by Eq. 16-19,

$$\begin{aligned} V_{BB} &= I_B R_B + V_{BE} + (I_B + I_C) R_E \\ &= 0.004 \times 45 + 0.7 + (0.204)15 \\ &= 0.18 + 0.7 + 3.06 \cong 4 \text{ V} \end{aligned}$$

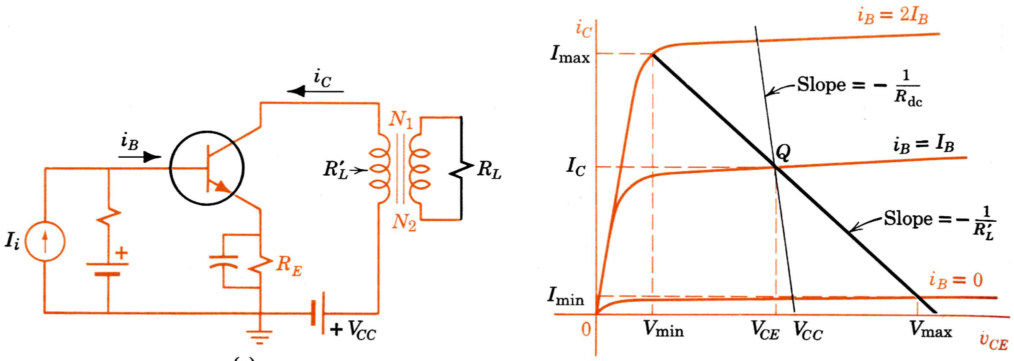
To complete the bias design, by Eqs. 16-20,

$$R_2 = R_B \frac{V_{CC}}{V_{BB}} = 45 \frac{58}{4} = 652 \cong 680\ \Omega$$

$$R_1 = \frac{R_2 R_B}{R_2 - R_B} = \frac{680 \times 45}{680 - 45} = 48 \cong 47\ \Omega$$

R_C instead of being transferred to R_L. The necessary supply voltage can be reduced and the power transfer improved by using a *transformer* to couple the output signal to the load.

As pointed out on p. 224, a transformer passes on ac signals but, since only *changing* magnetic flux induces secondary voltage, dc currents flow only in the low-resistance primary winding. With transformer coupling (Fig. 16.18), the supply voltage V_{CC} is approximately equal to V_{CE} instead of $V_{CE} + I_C R_C$; the dc power is correspondingly reduced. Also, by proper choice of the turn ratio a transformer permits impedance matching for maximum power transfer. For example, the low resistance of a loudspeaker (R_L) can be made to appear to a transistor as a much higher value (R'_L). For resistances, Eq. 8-28 becomes



(a)

(b)

Figure 16.18 Simplified circuit of a transformer-coupled power amplifier.

$$R'_L = \left(\frac{N_1}{N_2}\right)^2 R_L \tag{16-23}$$

To gain these benefits, the output of an audiofrequency amplifier is usually coupled to the load by a transformer. The performance of a well-designed transformer approaches the ideal over the major part of the audible range, but the output falls off at very low and very high frequencies. (See Fig. 22.20.)

Power and Efficiency

The calculation of output power and efficiency can be illustrated for the case of the simplified transformer-coupled transistor amplifier of Fig. 16.18. Here the chosen Q point and desired signal swing define the ac load line whose slope is $-1/R'_L$. Once R'_L is determined, the turns ratio N_1/N_2 is selected to provide a match with the actual load resistance R_L. The dc load line has a slope of $-1/R_{dc}$; in this case R_{dc} is just R_E plus the small winding resistance of the transformer primary, $I_C R_{dc} \ll V_{CE}$, and therefore $V_{CC} \cong V_{CE}$.

Assuming a sinusoidal signal current in a resistive load, the amplitude of the sinusoid is just one half the difference between the maximum and minimum values of current, and the rms value is

$$I_c = \frac{1}{2\sqrt{2}}(I_{\text{max}} - I_{\text{min}}) \tag{16-24}$$

The output signal power is

$$P_o = I_c^2 R_L' = V_c I_c = \frac{V_c^2}{R_L'} \tag{16-25}$$

where V_c is the rms value of the signal voltage across R_L'.

With transformer coupling, the average dc power supplied by the collector battery (R_E is to be small and is neglected for the moment) is

$$P_{CC} = V_{CC} I_C \cong V_{CE} I_C \tag{16-26}$$

For sinusoidal signals, the average power dissipated in the transistor is

$$\begin{aligned} P_D &= \frac{1}{2\pi}\int_0^{2\pi} v_{CE} i_C \, d(\omega t) \\ &= \frac{1}{2\pi}\int_0^{2\pi} (V_{CE} - \sqrt{2}V_c \sin \omega t)(I_C + \sqrt{2}I_c \sin \omega t)\, d(\omega t) \\ &= V_{CE} I_C - V_c I_c \end{aligned} \tag{16-27}$$

This interesting result indicates that the power dissipated is the difference between a constant input $V_{CE}I_C$ and a variable output $V_c I_c$. The losses to be dissipated are low when the signal output is high, and the power to be dissipated is maximum in the quiescent condition with no signal applied.

The efficiency of the output circuit is defined as signal power output over dc power input or

$$\text{Efficiency} = \frac{P_o}{P_I} = \frac{V_c I_c}{V_{CE} I_C} \tag{16-28}$$

The theoretical limit for efficiency in this type of amplifier can be determined by considering the idealized current amplifier whose characteristics are shown in Fig. 16.19. A signal of amplitude equal to I_B produces a total current swing from $2I_C$ to zero and a total voltage swing from $2V_{CE}$ to zero. The ideal efficiency is

$$\frac{P_o}{P_I} = \frac{(V_{CE}/\sqrt{2})(I_C/\sqrt{2})}{V_{CE}I_C} \times 100 = 50\% \tag{16-29}$$

At maximum signal input, the actual efficiency of transistor amplifiers is around 40 to 45%. The *average* efficiency with variable signal input is usually much lower. (See Example 5 on the next page.)

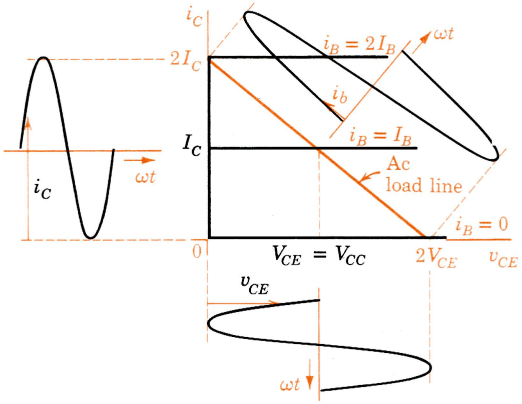


Figure 16.19 Operation of an idealized current amplifier.

Example 5

Modify the design of Example 4 to provide transformer coupling for maximum power output to a load of 5 Ω, and predict the amplifier output and efficiency.

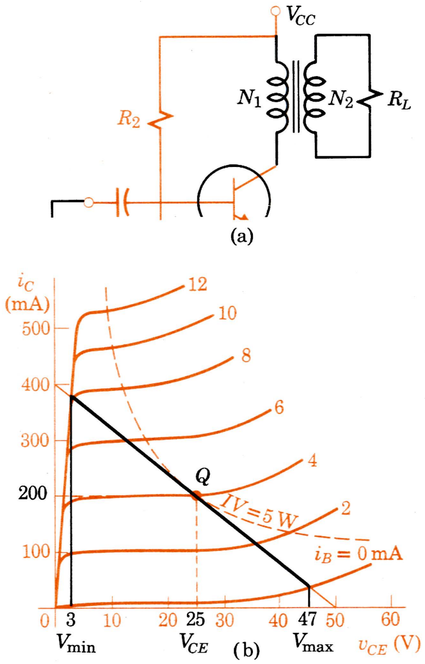


Figure 16.20 Amplifier performance.

The circuit modification is shown in Fig. 16.20a. The transformer should match the 5-Ω load to the 125-Ω load line and a commercial unit would be so labeled. The turns ratio should be (by Eq. 16-23)

$$N_1/N_2 = \sqrt{R_L'/R_L} = \sqrt{125/5} = 5$$

Now the necessary supply voltage is only

$$V_{CC} = V_{CE} + 3 = 25 + 3 = 28\text{ V}$$

(R_1 and R_2 must be recalculated.)

To determine the amplifier performance, we note that the ac load line in Fig. 16.20b intersects the $i_B = 8$ mA line at $V_{\min} \cong 3$ V, and the $i_B = 0$ line at $V_{\max} = 47$ V; therefore the maximum undistorted voltage swing is

$$\frac{V_{\max} - V_{\min}}{2} = \frac{47 - 3}{2} = 22\text{ V} = \sqrt{2}V_c$$

The power output is

$$P_o = \frac{V_c^2}{R_L'} = \frac{(22/\sqrt{2})^2}{125} = 1.94\text{ W}$$

Neglecting R_E, the collector circuit efficiency is

$$\frac{P_o}{P_I} = \frac{1.94}{5} \times 100 \cong 39\%$$

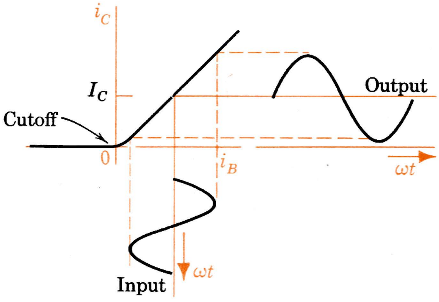
If the emitter resistance loss is included, $V_{CC} = 28$ V and, assuming $I_E \cong I_C$, the efficiency is

$$\frac{P_o}{V_{CC}I_C} = \frac{1.94}{28 \times 0.2} \times 100 \cong 35\%$$

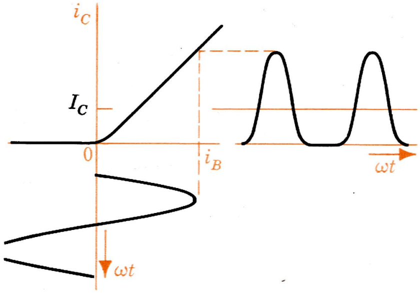
A small but significant amount of input power is required to "drive" the base of a large-signal amplifier. In Example 5, the rms value of base signal current is $4/\sqrt{2}$ mA and the necessary rms value of base-emitter signal voltage may be around $0.5/\sqrt{2}$ V. (The actual value can be predicted from I_C vs V_{BE} characteristics provided by the manufacturer.) The required input power of approximately 1 mW is supplied by a small-signal amplifier of the type discussed in Chapter 18.

Other Types of Amplifiers

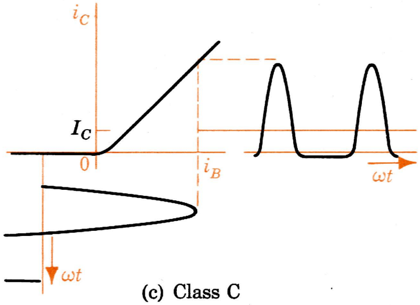
The emphasis in this chapter is on untuned power amplifiers. Some special forms of these amplifiers and some different types of amplifiers deserve mention.



(a) Class A



(b) Class B



(c) Class C

Figure 16.21 Amplifier operation modes.

Class B and C Operation

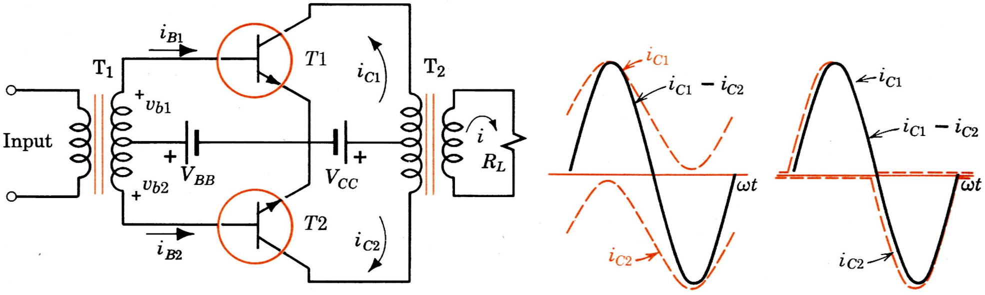
If corresponding values of i_B and i_C are obtained from the ac load line of Fig. 16.16b and plotted as in Fig. 16.21, the transfer characteristic is obtained. For distortion-free operation of untuned amplifiers the operating point is placed in the center of the linear part of the transfer characteristic and output current flows throughout the input-signal cycle. This condition is called *class* A operation and results in low distortion and also low efficiency because of the high value of I_C with respect to the output signal current.

If the amplifier is biased to cutoff, output current flows only during the positive half-cycle and the output is badly distorted. The resulting *class B* operation is useful in a push-pull circuit (Fig. 16.22). The second transistor supplies the other half-cycle of output current and the even harmonic distortion is cancelled. The average value of collector current is much lower and, therefore, the dc power input is less and the efficiency is higher.

If the amplifier is biased beyond cutoff, output current flows during only a small part of the cycle. With *class* C operation, the output is highly distorted, but with a tuned load impedance this presents no problem (Fig. 16.23). The tuned circuit selects the fundamental component of the signal and rejects all others. A sinusoidal output is obtained and the efficiency is high because the average collector current is relatively low.

Push-Pull Amplifiers

Much of the distortion introduced in large-signal amplifiers can be eliminated by using two transistors in the *push-pull* circuit of Fig. 16.22. The input transformer T_1 receives a sinusoidal voltage from a low-level source. The signals applied to the two transistors are 180° out of phase, and the resulting collector currents are 180° out of phase. The output transformer T_2 delivers to the load a current that is proportional to the difference of the two collector currents. Any even-harmonic distortion components tend to cancel and the only distortion is that due to odd harmonics. Also the performance of transformer T_2 (a critical component) is improved because the dc components of i_{C1} and i_{C2} just cancel and magnetic core saturation and the accompanying nonlinearity are avoided.



(a) Circuit diagram (b) Class A operation (c) Class B operation

Figure 16.22 Basic push-pull amplifier circuit.

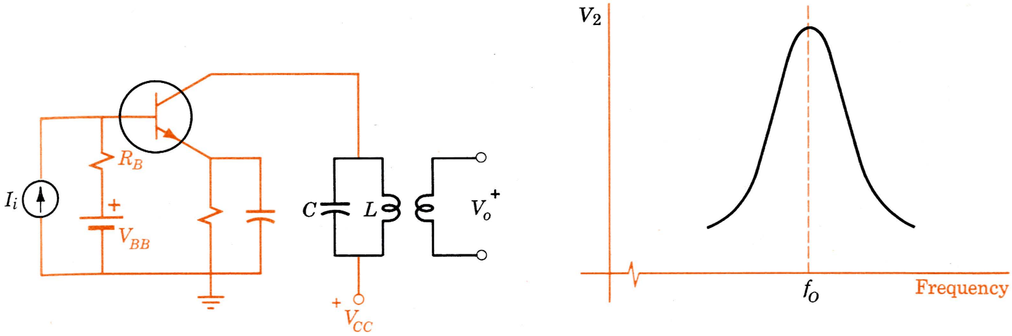
The push-pull circuit is particularly useful for class B operation. Bias supply V_{BB} is set for the turn-on voltage, about 0.7 V for silicon transistors. (In practice, V_{BB} is obtained from a voltage divider across V_{CC} or from the voltage drop across a diode.) As v_{B1} goes positive, i_{B1} begins to flow and i_{C1} follows (Fig. 16.22c). Since v_{B2} is negative, transistor T2 is cut off. When v_{B1} goes negative, v_{B2} is positive and transistor T2 takes over. If the two transistors have identical characteristics, as they will have in integrated circuit form, even harmonics are canceled and the result is a nearly pure sinusoidal output.

Following reasoning similar to that in the derivation of Eq. 16-29, the ideal class B efficiency can be calculated. For two transistors, $P_I = 2V_{CC}I_C$ and $I_C = I_{dc} = I_{max}/\pi$. Therefore,

$$\text{Efficiency} = \frac{P_o}{P_I} = \frac{V_c I_c}{2V_{CC}I_C} = \frac{(V_{max}/\sqrt{2})(I_{max}/\sqrt{2})}{2V_{max}(I_{max}/\pi)}$$

$$= \frac{\pi}{4} = 0.785 \quad \text{or} \quad 78.5\%$$

Since the *losses* in a transistor cannot exceed the rated heat dissipation, two transistors in class B can supply nearly six times the power output of one similar transistor in class A. (See Exercise 24.)



(a) Circuit

(b) Frequency response

Figure 16.23 A transistor tuned amplifier.

Tuned Amplifiers

The gain of a transistor amplifier depends on the load impedance. If a high-Q parallel resonant circuit is used for the load (Fig. 16.23), a very high resistance is presented at the resonant frequency and therefore the voltage gain is high. The frequency-response curve has the same shape as that of the resonant circuit. A narrow band of frequencies near resonance is amplified well, but signals removed from the resonant frequency are discriminated against.

The high selectivity of the load impedance eliminates nonlinear distortion; any harmonics in the input signal or in the collector current itself develop little voltage across the load impedance. As long as the collector current has a component at the resonant frequency, the output is nearly sinusoidal. With distortion eliminated, high efficiency is achieved by operating the transistor in a nonlinear region.

SUMMARY

- An amplifier is a device for raising the level of a signal voltage, current, or power. For sinusoidal signals, the voltage gain is

$$\mathbf{A} = \frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = A\, e^{j\theta}$$

If A and θ are constant (or if $\theta = k\omega$), the amplification is linear and there is no distortion.

- In practical amplifiers, power supplies and biasing networks must maintain operation at the proper point, coupling circuits must transfer signals from one stage to the next without excessive discrimination, and load impedances must provide desired output without requiring excessive supply voltages.

- For FETs used as small-signal amplifiers, biasing is not critical and various circuits are available.
For BJTs used as large-signal amplifiers, biasing by a combination of emitter resistor and voltage divider works satisfactorily.
For noncritical design, an approximate bias procedure is available.

- The permissible operating region of a transistor is defined by maximum allowable current, voltage, power, and distortion.
For large-signal class A operation, the midpoint of a properly located ac load line is the optimum quiescent point.

- Transformer coupling reduces the required supply voltage, isolates the signal output, permits impedance matching, and improves efficiency.
In an idealized transformer-coupled, class A current amplifier:

$$\begin{aligned}\text{Input power (dc)} &= P_I = V_{CC}I_C = V_{CE}I_C \\ \text{Output power (max)} &= P_o = V_cI_c = \tfrac{1}{2}V_{CE}I_C = \tfrac{1}{2}P_I\end{aligned}$$

- Class B and C amplifiers operate efficiently because output current is cut off during ineffective portions of the signal cycle.
The push-pull amplifier reduces distortion for a given output and improves the operation of output transformers.
The tuned amplifier uses a resonant load circuit to obtain high selectivity and low distortion.

REFERENCES

1. D. L. Schilling and C. Belove, *Electronic Circuits: Discrete and Integrated,* McGraw-Hill Book Co., New York, 1968, ch. 4 and 5.
2. J. Millman and C. C. Halkias, *Integrated Electronics,* McGraw-Hill Book Co., New York, 1972, ch. 9, 10, and 18.
3. A. P. Malvino, *Electronic Principles,* McGraw-Hill Book Co., New York, 1973, ch. 7, 10, and 13.

REVIEW QUESTIONS

1. Distinguish between tuned and wide-band amplifiers and give an application of each (other than those in the text).
2. Distinguish between frequency, phase, and amplitude distortion. Explain the effect of each in a hi-fi system.
3. What determines the dynamic range of an amplifier?
4. What is the purpose of biasing in a MOSFET? In a BJT?
5. What are the relative merits of RC and transformer coupling?
6. Explain the practical considerations involved in the selection of values for: R_1, R_2, R_D, R_S in Fig. 16.8b.
7. Explain the operation of drain-feedback bias.
8. Why is fixed-current transistor bias unsatisfactory?
9. Explain thermal runaway.
10. Explain the function of each circuit element in Fig. 16.24.
11. Explain how the circuit of Fig. 16.24 stabilizes I_C if V_{CE} changes.
12. What are the practical limits on R_1 and R_3 in Fig. 16.24.
13. How are the quiescent point and load line related to the permissible operating region?
14. Sketch a set of transistor characteristics and indicate regions of high amplitude distortion.
15. Explain the statement: "A power transistor runs cool when the power output is high."
16. How does a transformer improve amplifier efficiency?
17. Why is impedance matching desirable?
18. Explain how push-pull amplifiers work. Where are they used?
19. How are "tuned amplifiers" tuned? Where are they used?
20. Explain why class B operation is more efficient than class A.

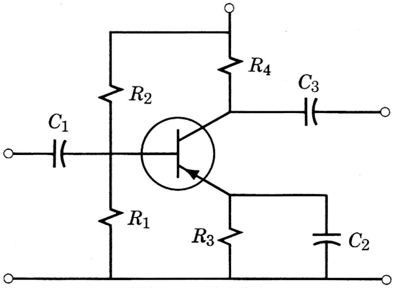


Figure 16.24

EXERCISES

1. Explain by means of sketches the differences between amplitude, frequency, and phase distortion. How would each be evidenced in a hi-fi system?
2. A MOSFET with the characteristics of Fig. 16.5a is used as a voltage amplifier with $V_{DD} = 20$ V and $R_D = 6$ kΩ.
   (a) Draw a "zero-bias" circuit and predict the quiescent values I_D and V_{DS}.
   (b) Sketch the transfer characteristic for a "zero-bias" circuit.
   (c) Estimate from the transfer characteristic, the output signal current i_d for an input voltage $v_i = 0.5 \sin \omega t$ V.
   (d) Estimate the voltage gain of the amplifier.
3. Redraw the circuit of Fig. 16.5c modified by inserting resistance R_S between the S terminal and ground.
   (a) Derive an equation for i_D as $f(v_{DS})$ for the modified circuit.
   (b) The MOSFET of Fig. 16.5a is to operate at $I_D = 1$ mA and $V_{DS} = 8$ V; for $R_D = 5$ kΩ, specify R_S and V_{DD}. Check your answers by drawing the load line.
4. A JFET for which $I_{DSS} = 8$ mA and $V_p = -6$ V is used in the circuit of Fig. 16.25 where $V_{DD} = 32$ V, $V_{GG} = 10$ V, $R_D = 3$ kΩ, $R_S = 2.5$ kΩ, and $R_G = 10$ MΩ.
   (a) Estimate I_D by assuming $V_{GS} = 0$ and $I_G = 0$.
   (b) Using the estimated value of I_D, estimate V_{GS} by Eq. 13-1 and recalculate I_D from the circuit. Draw a conclusion about the method in (a).

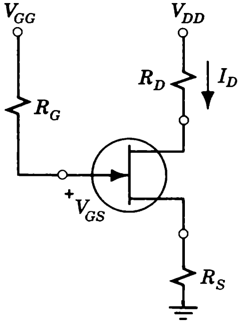


Figure 16.25

5. In the circuit of Fig. 16.7, $V_{DD} = 24$ V, $R_D = 3$ kΩ, $R_S = 5$ kΩ, $R_1 = 6$ MΩ, and $R_2 = 12$ MΩ.
   (a) Assume $V_{GS} = -2$ V and predict the quiescent values I_D and V_{DS}.
   (b) Assume V_{GS} is actually -1 V and repeat (a); draw a conclusion.
6. A JFET for which $I_{DSS} = 5$ mA and $V_{GS(\mathrm{OFF})} = -6$ V is to be operated at a Q point defined by $I_D = 3$ mA and $V_{DS} = 10$ V. Assuming $R_{\mathrm{in}} \geq 10$ MΩ and $R_D = 3$ kΩ, design an appropriate bias circuit.
7. The enhancement MOSFET of Fig. 13.3 is to be operated in the normal region at $I_D = -2$ mA. Draw a drain-feedback bias circuit and specify V_{DD} assuming $R_D = 7$ kΩ.
8. The data sheet of an enhancement MOSFET specifies $V_T = -3$ V and $I_D = -4$ mA for $V_{GS} = V_{DS} = -8$ V.
   (a) Determine V_{GS} for $I_D = -2$ mA.
   (b) Specify R_D for drain-feedback bias at $I_D = -2$ mA with $V_{DD} = 16$ V.
9. In the circuit of Fig. 16.9a, $I_B = 0.1$ mA, I_{CBO} for the germanium transistor is 5 μA, and β ranges from 50 to 200.
   (a) Predict the corresponding range of values of I_C.
   (b) If $\beta = 150$ and if I_{CBO} quadruples as a result of temperature increase, predict the change in I_C.
10. A silicon transistor with a nominal $\beta = 100$ is to operate with $I_C \cong 2$ mA and $V_{CE} = 6$ V in the circuit of Fig. 16.9.
   (a) For $V_{CC} = 12$ V, specify R_C and R_B.
   (b) If the β is actually 40, estimate the actual value of I_C for the value of R_B specified in part (a).
11. The transistor of Exercise 10 is to operate under the same quiescent conditions in the circuit of Fig. 16.10.
   (a) For $R_C = 3$ kΩ and $R_E = 0.5$ kΩ, specify V_{CC}.
   (b) For $R_1 = 12$ kΩ and $R_2 = 68$ kΩ, calculate V_{BB} and R_B.
   (c) If the β is actually 40, estimate the actual value of I_C.
   (d) Compare the performance of the circuit with that in Exercise 10.
12. In the self-bias circuit (Fig. 16.10) $R_1 = 70$ kΩ, $R_2 = 140$ kΩ, $R_E = 3.9$ kΩ, and $V_{CC} = 15$ V. Estimate I_C and specify resistor R_C to establish the quiescent voltage across the silicon transistor (V_{CE}) at 5 V.
13. A silicon transistor whose β may vary from 50 to 200 (nominal $\beta = 100$) is to be operated at $I_C = 2$ mA and $V_{CE} = 5$ V. For this amplifier application, $R_C = 2$ kΩ.
   (a) Using the approximate procedure, design an appropriate bias network and draw a labeled circuit diagram.
   (b) Determine the maximum variation in I_C expected for the given variation in β.
14. A germanium transistor with a nominal β of 60 is to be operated at $I_C = 0.5$ mA and $V_{CE} = 4$ V with the available $V_{CC} = 10$ V.
   (a) Using the approximate procedure, design an appropriate self-biasing network and draw a labeled circuit diagram.
   (b) If the actual β varies from 20 to 100, predict the maximum variation in I_C.
15. The data sheet for a 2N3114 specifies a maximum power dissipation of 5 W with the case at 25°C and a "derating factor" of 28.6 mW/°C above 25°C to maintain a maximum junction temperature of 200°C.
   (a) Explain the term "derating factor" in terms of energy transformations within the transistor and between the transistor and the case.
   (b) If during operation the case temperature of the 2N3114 may rise to 75°C, estimate the allowable power dissipation.

16. The collector dissipation for the transistor of Fig. 16.20b is to be limited to 3 W, maximum I_C is 1 A, and maximum V_{CE} is to be 60 V.
   (a) Reproduce the characteristics and sketch in the permissible operating region.
   (b) For operation at $V_{CE} = 20$ V, specify a suitable R_{ac} and the nominal I_B for maximum power output.

17. A silicon transistor with a β of 50 and a maximum allowable collector dissipation of 10 W is to be used in the circuit of Fig. 16.15a. The maximum collector current is 2 A and the maximum collector voltage is 50 V.
   (a) Making and stating any desirable simplifying assumptions, sketch the permissible operating region for this transistor.
   (b) Estimate the maximum possible signal power output in R_L and the efficiency.

18. The silicon transistor in the circuit of Fig. 16.16(a) has a nominal β of 100; $V_{CC} = 15$ V, $R_1 = 10$ kΩ, and $R_2 = 30$ kΩ.
   (a) Specify R_E and R_C to put the Q point at $V_{CE} = 6$ V and $I_C = 2$ mA.
   (b) Sketch the i_C vs v_{CE} characteristics and draw the dc load line.
   (c) For $R_L = 3$ kΩ, draw the ac load line and estimate the output voltage v_o for an input current $i_b = 10 \sin \omega t$ μA.

19. Repeat part (b) of Exercise 17 assuming that R_L is replaced by an ideal transformer providing matched coupling to $R_L = 750$ Ω. Specify the turn ratio of the transformer. Compare the results to those for Exercise 17, and draw a conclusion.

20. A transistor is rated at $I_C(\max) = 1$ A, $V_{CE}(\max) = 60$ V, $P_D(\max) = 3$ W. Estimate the maximum signal power output for an amplifier using this transistor in class A operation.

21. The permissible operating region of a power transistor is defined by $P_D = 5$ W, $I_C(\max) = 1$ A, $V_{CE}(\max) = 100$ V, and $V_{CE}(\min) = 2$ V.
   (a) Sketch the transistor characteristics and select an appropriate quiescent point for operation in the circuit of Fig. 16.15a.
   (b) Specify R_L for nearly maximum power output.
   (c) Calculate total dc power in, maximum signal power out, and overall efficiency.
   (d) If the actual load is 10 Ω, specify the appropriate transformer turn ratio and recalculate the overall efficiency of the amplifier.

22. A transistor rated at $I_C(\max) = 1.6$ A, $V_{CE}(\max) = 50$ V, $V_{CE(\text{sat})} \cong 0$ V, $I_{CEO} \cong 0$, and $P_D(\max) = 10$ W is used in the circuit of Fig. 16.15a.
   (a) Draw, approximately to scale, the permissible operating region.
   (b) For $R_L = 100$ Ω, select the "best" quiescent point, explaining your reasoning.
   (c) Predict the average collector dissipation, the maximum power output to R_L, and the collector circuit efficiency.
   (d) If your Q point does not take full advantage of the capability of the transistor, rearrange the circuit to provide maximum possible power to the same 100-Ω load and repeat part (c).

23. If two transistors of the type described in Exercise 20 are used in a class B push-pull amplifier, estimate the maximum signal power output and draw a conclusion.

24. For Class B operation,
   (a) Show the operation of an amplifier on the idealized characteristics of Fig. 16.19.
   (b) Demonstrate that two transistors in class B can supply nearly six times the power of one similar transistor in class A.

PROBLEMS

1. An inexperienced designer uses the transistor of Example 2 in the circuit of Fig. 16.9(a) with $V_{CC} = 20$ V and $R_C = 5$ kΩ.
   (a) Specify R_B for a nominal $I_C = 1$ mA.
   (b) Predict the extreme variation in I_C for this circuit, compare to the variation in Example 2, and draw a conclusion regarding the effectiveness of this method.

2. A JFET (somewhat like Fig. 16.6a) and a BJT (somewhat like Fig. 16.9b) are connected in the *current-source bias* circuit of Fig. 16.26. Estimate I_D and V_{DS}. (*Hint:* Estimate the voltages at points 1, 2, and 3 in turn.) Evaluate the stability of this bias circuit.

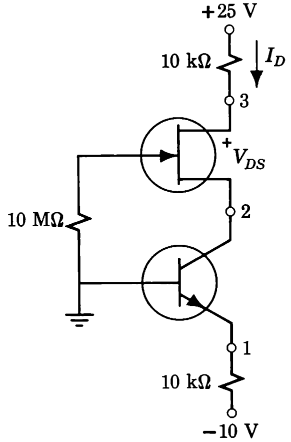


Figure 16.26

3. A 2N3114 transistor (Fig. A3) is to be operated as a single-stage audiofrequency amplifier at $I_C = 2$ mA and $V_{CE} = 40$ V with $R_C = 10$ kΩ. Draw an appropriate wiring diagram and, stating all assumptions, design the bias network.
4. The circuit of Fig. 16.27 is suggested as a simple means for obtaining improved stability.
   (a) Explain qualitatively what happens if I_C tends to rise as a result of an increase in I_{CBO} or β.
   (b) Derive an approximate expression for I_C in terms of I_{CBO} and β (somewhat like Eq. 16.10).
   (c) Under what conditions (i.e., for what relation between R_C and R_B) will I_C be insensitive to changes in β?
5. Prove that the point at which a load line is tangent to the hyperbola $P_D = V_{CE}I_{CE}$ is the midpoint of the load line.
6. A transformer-coupled audio amplifier is to supply 1 W to a 16-Ω loudspeaker. Available is a transistor whose permissible operating range is defined by $P_D = 4$ W, V_{CE} (max) $=$ 80 V, V_{CE} (min) $= 2$ V at $I_C = 200$ mA, and I_C (max) $= 1$ A. *Stating* all assumptions, design the amplifier (draw the wiring diagram and specify the quiescent point, bias resistances, supply voltage, capacitor values, and transformer turns ratio) and estimate the total dc power requirement.
7. Demonstrate that the maximum theoretical efficiency for a class A amplifier with a resistive load in the collector circuit (not transformer coupled) is 25%.
8. The circuit of Fig. 16.28 has two outputs labeled 1 and 2. Predict I_C and V_{CE} for $v_i = 0$. Predict and sketch the output voltages for $v_i = V_m \sin \omega t$. What function does this circuit perform? Where would it be useful?

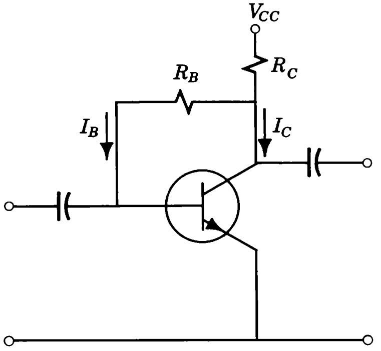


Figure 16.27

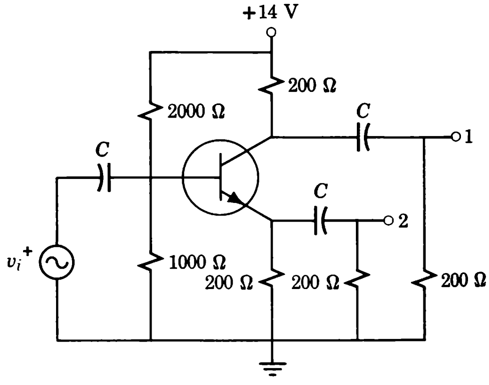


Figure 16.28

Chapter 17

Small-Signal Models

Large and Small Signals
Diodes
Field-Effect Transistors
Junction Transistors

The use of models to represent complicated phenomena or devices or systems is common practice throughout engineering and science. On the basis of experimental observations, the scientist may propose a tentative model that is then tested under various conditions. An adequately tested hypothesis is called a *theory;* a repeatedly confirmed theory becomes a *law.* The engineer uses models to simplify the analysis of a known device or system or to predict the behavior of a proposed design. These models may be physical or mathematical in form; a linear model amenable to mathematical analysis is particularly desirable.

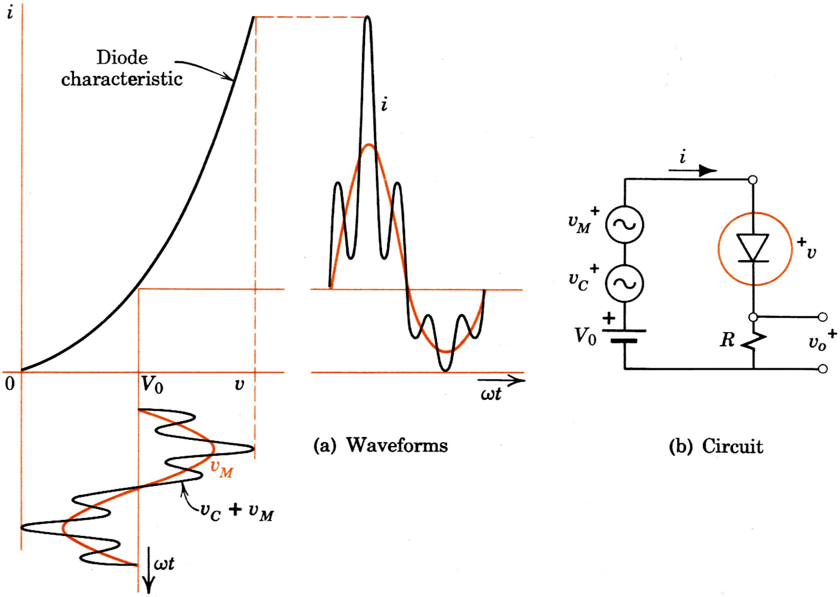
The linear electrical model of a device is frequently a *circuit model* composed of ideal circuit elements. By using familiar basic elements, it is possible to devise relatively simple circuit models to represent complicated physical devices under specified conditions. The models are derived from the characteristic curves of the real devices and, when properly devised, can be used to predict accurately the behavior of semiconductor devices (or vacuum tubes) in practical applications such as voltage and current amplification. More sophisticated models are used in computer-aided design for optimum performance. The character of the model depends on the size of the signals expected.

Large and Small Signals

The circuit models derived in Chapters 12 and 14 are based on piecewise linearization of the characteristic curves and are useful for predicting the behavior of diodes and transistors under static conditions or where relatively large swings in voltage occur. A different approach is necessary if we are interested in the response to small signals superimposed upon dc values.

A Mathematical Model

Consider the circuit of Fig. 17.1 in which two sinusoidal signals $v_M(t)$ and $v_C(t)$ are superimposed on a bias voltage V_0 and the total is applied to a diode circuit.



(a) Waveforms (b) Circuit

Figure 17.1 Effect of nonlinearity on the sum of two signals.

Using the analytical technique of Chapter 2, we can represent a portion of the diode characteristic by the power series

$$i = a_0 + a_1 v + a_2 v^2 \tag{17-1}$$

If the voltage drop across R is relatively small, the full voltage appears across the diode and

$$v = V_0 + v_M + v_C = V_0 + V_M \sin \omega_M t + V_C \sin \omega_C t \tag{17-2}$$

Then

$$\begin{aligned} i = a_0 &+ a_1 V_0 + a_1 V_M \sin \omega_M t + a_1 V_C \sin \omega_C t \\ &+ a_2 V_0^2 + a_2 V_M^2 \sin^2 \omega_M t + a_2 V_C^2 \sin^2 \omega_C t \\ &+ 2a_2 V_0 V_M \sin \omega_M t + 2a_2 V_0 V_C \sin \omega_C t \\ &+ 2a_2 V_M V_C \sin \omega_M t \sin \omega_C t \end{aligned}$$

Substituting for $\sin^2 \omega t$ and $\sin \omega_M t \sin \omega_C t$ and rearranging terms,

$$\begin{aligned} i = a_0 &+ a_1 V_0 + a_2 V_0^2 + \tfrac{1}{2} a_2 V_M^2 + \tfrac{1}{2} a_2 V_C^2 \\ &+ (a_1 + 2a_2 V_0) V_M \sin \omega_M t + (a_1 + 2a_2 V_0) V_C \sin \omega_C t \\ &- \tfrac{1}{2} a_2 V_M^2 \cos 2\omega_M t - \tfrac{1}{2} a_2 V_C^2 \cos 2\omega_C t \\ &+ a_2 V_M V_C \cos (\omega_C - \omega_M)t - a_2 V_M V_C \cos (\omega_C + \omega_M)t \qquad (17\text{-}3) \end{aligned}$$

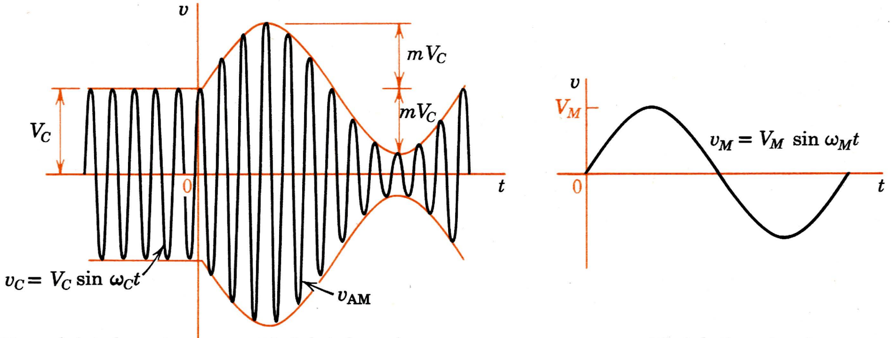
The output voltage v_o is directly proportional to i and contains the same eleven components. The first five terms constitute a dc component dependent on the amplitude of the input signals, as expected in a nonlinear device. The next two terms are replicas of the input signals resulting from the linear term of Eq. 17-1. Then there are two terms representing second harmonics generated by the nonlinearity. The last two terms are of particular interest in communication.

Amplitude Modulation

In a typical communication or control system, a high-frequency easily propagated signal serves as a *carrier* and information is superimposed on the carrier in the process called *modulation*. In amplitude modulation (AM), the amplitude of the carrier $v_C = V_C \sin \omega_C t$ is varied by the modulating signal $v_M = V_M \sin \omega_M t$. The equation of the amplitude-modulated wave is

$$v_{\mathrm{AM}} = (1 + m \sin \omega_M t) V_C \sin \omega_C t \qquad (17\text{-}4)$$

where m is the *degree of modulation* $= V_M/V_C$, approximately 80% in Fig. 17.2.



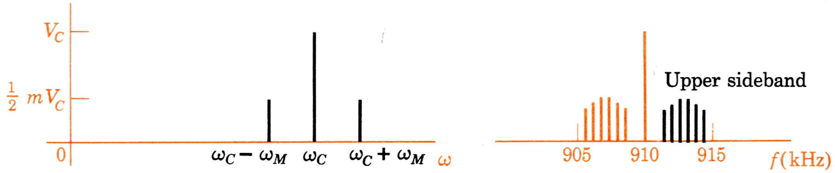
Unmodulated carrier **Modulated carrier** **Modulating signal**

Figure 17.2 A sinusoidal carrier amplitude-modulated by a sinusoidal signal.

Expanding Eq. 17-4,

$$\begin{aligned} v_{\mathrm{AM}} &= V_C \sin \omega_C t + m V_C \sin \omega_M t \sin \omega_C t \\ &= V_C \sin \omega_C t + \tfrac{1}{2} m V_C \cos (\omega_C - \omega_M)t - \tfrac{1}{2} m V_C \cos (\omega_C + \omega_M)t \qquad (17\text{-}5) \end{aligned}$$

This equation reveals that the amplitude-modulated wave consists of three sinusoidal components of constant amplitude. The carrier and the upper and lower *side frequencies* are shown in the frequency spectrum of Fig. 17.3a. In general the modulating signal is a complicated wave containing several components and the result is a carrier plus upper and lower *sidebands.* If the 910 kHz carrier of an AM station is modulated by frequencies from 100 to 4500 Hz, the frequency spectrum is as shown in Fig. 17.3b. The station is assigned a *channel* 10 kHz wide to accommodate side frequencies from 0 to 5000 Hz. The amplifiers in AM transmitters and receivers are designed to amplify well over a 10-kHz bandwidth and to discriminate against frequencies outside this range. In comparison, channels for frequency modulation (FM) broadcast are 200 kHz wide and permit lower noise and higher frequency response for greater fidelity. For satisfactory television transmission, sidebands approximately 6 MHz wide are required.



(a) (b)

Figure 17.3 Frequency spectra of AM waves.

Example 1

A carrier signal $v_C = 100 \sin 10^7 t$ V is to be modulated by a signal $v_M = 100 \sin 10^4 t$ V, using a device with a characteristic defined by $i = v + 0.003v^2$ mA. The resulting current is filtered by a circuit that passes all frequencies from 9.9 to 10.1 Mrad/sec and rejects all others. Describe the output signal.

Comparing the device characteristic to Eq. 17-1, $a_0 = 0$, $a_1 = 1$, and $a_2 = 0.003$. Only the seventh, tenth, and eleventh components in Eq. 17-3 have frequencies within the passband of the filter.

The magnitudes of the carrier and side-frequency components are:

$$a_1 V_C = 100 \text{ mA} \quad \text{and} \quad a_2 V_M V_C = 30 \text{ mA}$$

The output, an amplitude-modulated signal, is

$$i = 100 \sin 10^7 t - 30 \cos 1.001 \times 10^7 t + 30 \cos 0.999 \times 10^7 t \text{ mA}$$

Comparing the second term with the corresponding term in Eq. 17-5,

$$\tfrac{1}{2} m I_C = \tfrac{1}{2} m \times 100 = 30 \quad \text{or} \quad m = \frac{60}{100}$$

and the degree of modulation is 60%.

Effect of Signal Size

As indicated in Example 1, a diode can be used as a *modulator.* The same nonlinear device will also serve as a *detector* to recover the information from the three radiofrequency signals, or as a *mixer* to obtain the difference or "beat frequency" between two sinusoidal signals. (See Problems 1 and 3.) We know that the diode can also be used as a switch or as a rectifier. In all these functions, the signal excursions extend into the nonlinear regions of the diode characteristics. Obviously, the function actually performed is determined by the relative size of the coefficients of the various components in the output. These in turn depend on the factors a_0, a_1, and a_2 in the device characteristic, the bias voltage V_0, and the signal amplitudes V_M and V_C.

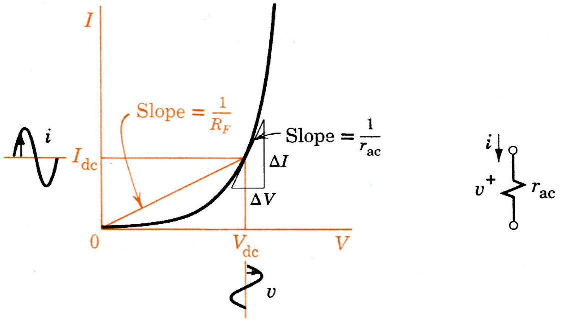
For linear applications, where the distorting effects of nonlinearity are to be avoided, the bias voltage is selected to place operation on a portion of the characteristic where factor a_2 is relatively small. Then the signal amplitudes are kept small. Since the coefficients of the unwanted components vary as the square or the product of the signal amplitudes, any device can be considered to be linear if the signals are sufficiently small. By definition, a "small signal" is one to which a given device responds in a linear fashion.

Diodes

If we are interested in the response of a diode to small signals superimposed upon dc values, a special circuit model is necessary.

Dynamic Resistance

As shown in Fig. 17.4, a small signal voltage v superimposed on a steady voltage V_{dc} produces a corresponding signal current i. The current-voltage relation for small signals is defined by the slope of the curve at the operating point specified by V_{dc}. The *dynamic resistance* or *ac resistance* is defined as



(a) Characteristic (b) Circuit model

Figure 17.4 Dynamic resistance of a diode.

$$r_{ac} = \frac{dv}{di} \cong \frac{\Delta v}{\Delta i} \quad \text{in ohms} \tag{17-6}$$

In general, the dynamic resistance r_{ac} will differ appreciably from the static forward resistance defined as $R_F = V_{\mathrm{dc}}/I_{\mathrm{dc}}$.

The value of r_{ac} can be predicted for a theoretical diode for which

$$I = I_s(e^{eV/kT} - 1)$$

as in Eq. 11-21. Taking the derivative,

$$\frac{dI}{dV} = \frac{e}{kT} I_s\, e^{eV/kT} = \frac{e}{kT}(I + I_s) \cong \frac{eI}{kT} \tag{17-7}$$

Hence, for this theoretical diode,

$$r_{\mathrm{ac}} = \frac{dV}{dI} = \frac{v}{i} = \frac{kT}{eI} = \frac{kT/e}{I} \tag{17-8}$$

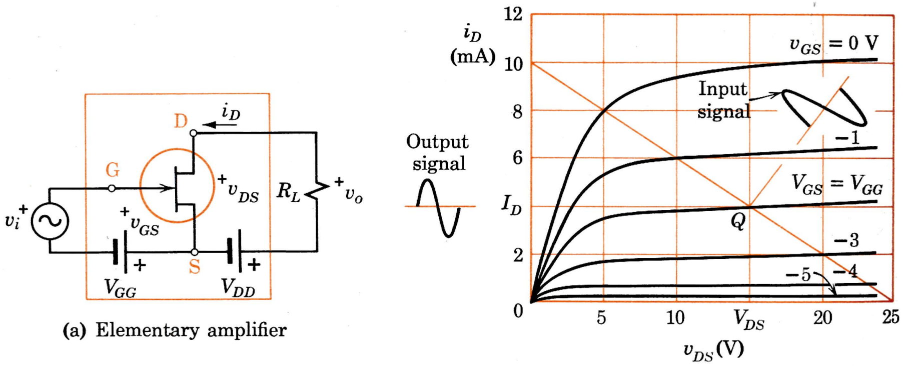
The dynamic resistance is directly proportional to absolute temperature T and inversely proportional to dc diode current I. At room temperature, $T = 293\ \mathrm{K} \cong 300\ \mathrm{K}$, $kT/e \cong 0.025\ \mathrm{V} = 25\ \mathrm{mV}$. Then

$$r_{\mathrm{ac}} = \frac{0.025}{I} = \frac{25}{I(\text{in mA})}\ \text{ohms} \tag{17-9}$$

This provides a simple means for estimating r_{ac} in practical diodes; however, the ohmic body resistance of the semiconductor may be the dominant factor at large currents.

Field-Effect Transistors

In the elementary amplifier of Fig. 17.5a, the input is a signal voltage v_i superimposed on the gate-source bias voltage V_{GG}. The signal voltage v_{gs} modulates the width of the conducting channel and produces a signal component i_d



(a) Elementary amplifier

(b) Characteristics

Figure 17.5 Small-signal operation of an FET.

of the drain current i_D. The signal current develops across resistor R_L an amplified version of the input voltage.

The quantitative behavior of a small-signal amplifier could be determined graphically as we did in the case of large-signal BJT amplifiers, but this approach is not good for two reasons. First, the work is laborious and the results lack the generality needed in optimizing a design. Second, individual devices of a given type vary widely in their characteristics and the curves provided by the manufacturer are only general indications of how that type behaves on the average. A better approach is to devise a convenient model that will accurately predict device performance within limitations. If we limit our consideration to small signals superimposed on dc values, we are justified in assuming linear relations among signals in devices that are highly nonlinear in their large-signal behavior. On this basis, the complicated physical device inside the colored line of Fig. 17.5a is to be replaced, insofar as small signals are concerned, by a linear circuit model.

Two-Port Parameters

A diode is a two-terminal device or one-port, and its circuit model also has a single pair of terminals. Transistors are two-ports and can be represented as in Fig. 17.6. In Chapter 8 equations were derived for replacing a two-port network by a T or Π combination of one-port impedances. In deriving the circuit models of transistors, important differences result from the fact that these are active, nonlinear devices, not the passive, linear networks treated previously. However, the same basic approach can be used.

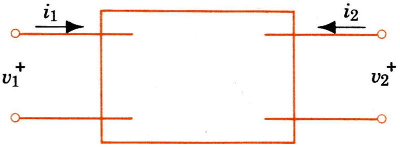


Figure 17.6 The general two-port.

In general, of the four variables identified only two are independent, and the behavior can be defined by a pair of equations such as

$$\begin{cases} v_1 = f_1(i_1, i_2) \\ v_2 = f_2(i_1, i_2) \end{cases} \quad \text{or} \quad \begin{cases} i_1 = f_3(v_1, v_2) \\ i_2 = f_4(v_1, v_2) \end{cases} \tag{17-10}$$

For linear networks, the first pair of equations would define a set of impedances as parameters† and the second pair would define a set of admittances as parameters. Another acceptable formulation is

$$\begin{cases} v_1 = f_5(i_1, v_2) \\ i_2 = f_6(i_1, v_2) \end{cases} \tag{17-11}$$

---

† This is more easily seen if the first pair of Eqs. 17-10 is rewritten:

$$\begin{cases} \mathbf{V}_1 = \mathbf{Z}_{11}\mathbf{I}_1 + \mathbf{Z}_{12}\mathbf{I}_2 \\ \mathbf{V}_2 = \mathbf{Z}_{21}\mathbf{I}_1 + \mathbf{Z}_{22}\mathbf{I}_2 \end{cases} \quad \text{for the linear sinusoidal case.}$$

For linear networks, the first equation of this pair would define an impedance and a voltage factor, and the second equation would define a current factor and an admittance. This mixture is called a set of *hybrid parameters*, and it is a particularly convenient set to use for electronic devices.

Evaluation of Parameters

In selecting an appropriate formulation for the field-effect transistor, we note that the input current is very small because in the JFET the gate-source junction is reverse biased and in the MOSFET the gate is insulated. Also, we note that the FET is a voltage-controlled device† (Fig. 17.5b). Therefore, we write (Eq. 17-10),

$$\begin{cases} i_G = 0 \\ i_D = f(v_{GS}, v_{DS}) \end{cases} \tag{17-12}$$

Since $i_G = 0$ at all times, only the second equation is significant. Because we are interested in small or "differential" changes in the variables, we express the total differential of the drain current as

$$di_D = \frac{\partial i_D}{\partial v_{GS}} dv_{GS} + \frac{\partial i_D}{\partial v_{DS}} dv_{DS} \tag{17-13}$$

Equation 17-13 says that a change in gate voltage dv_{GS} and a change in drain voltage dv_{DS} both contribute to a change in drain current di_D. The effect of the partial contributions is determined by the coefficients $\partial i_D/\partial v_{GS}$ and $\partial i_D/\partial v_{DS}$; these are the "partial derivatives" of the drain current with respect to gate voltage and drain voltage.

If the drain-source voltage is held constant, $dv_{DS} = 0$ and the partial derivative with respect to gate-source voltage is equal to the total derivative or

$$\frac{\partial i_D}{\partial v_{GS}} = \left.\frac{di_D}{dv_{GS}}\right|_{v_{DS}=k} = g_m \tag{17-14}$$

where g_m is the *transconductance* in mhos, so called because it is the ratio of a differential current at the output to the corresponding differential voltage at the input. Typical values of g_m lie between 500 and 10,000 $\mu\mho$.

A second FET parameter can be defined by holding the gate-source voltage constant so that $dv_{GS} = 0$ and

$$\frac{\partial i_D}{\partial v_{DS}} = \left.\frac{di_D}{dv_{DS}}\right|_{v_{GS}=k} = \frac{1}{r_d} \tag{17-15}$$

where r_d is the *dynamic drain resistance* and is just equal to the reciprocal of the slope of a line of constant v_{GS}. Typical values of r_d lie between 20 and 500 kΩ.

† The vacuum tube is also a "voltage-controlled device," and its small-signal model is similar to that for an FET. See pp. 413 ff of Reference 1 at the end of Ch. 16.

(Instead of r_d, the *output conductance* $g_{os} = g_d = 1/r_d$ may be specified.)

For the JFET or the DE MOSFET, the drain current in the normal operating region is (Eq. 13-1)

$$i_D = I_{DSS}\left(1 - \frac{v_{GS}}{V_p}\right)^2 \tag{17-16}$$

At a dc bias voltage V_{GS}, the transconductance is

$$g_m = \frac{di_D}{dv_{GS}} = -\frac{2I_{DSS}}{V_p}\left(1 - \frac{V_{GS}}{V_p}\right) = g_{mo}\left(1 - \frac{V_{GS}}{V_p}\right) = g_{mo}\sqrt{\frac{I_D}{I_{DSS}}} \tag{17-17}$$

The transconductance evaluated at $V_{GS} = 0$ is

$$g_{mo} = -\frac{2I_{DSS}}{V_p}\left(1 - \frac{0}{V_p}\right) = -\frac{2I_{DSS}}{V_p} \tag{17-18}$$

sometimes called y_{fs} the *forward transadmittance*. Since the manufacturer usually gives the value of g_{mo} (or y_{fs}) along with I_{DSS}, Eq. 17-17 can be used to calculate g_m for any operating value of I_D or V_{GS} (after solving for V_p from Eq. 17-18.) For E MOSFETs, the data sheet usually includes a curve of g_m vs I_D.

Example 2

Estimate the model parameters of the FET of Fig. 17.5b for operation at a Q point defined by $v_{DS} = +15$ V and $v_{GS} = -2$ V. Use Δi and Δv to approximate di and dv.

Taking increments from the given operating point and reading the corresponding values yields

$$g_m = \left.\frac{di_D}{dv_{GS}}\right|_{+15\text{V}} \cong \frac{\Delta i_D}{\Delta v_{GS}}$$

$$= \frac{(6.5 - 1.9)10^{-3}}{-1 + 3} = 2300\ \mu\mho$$

$$r_d = \left.\frac{1}{\dfrac{di_D}{dv_{DS}}}\right|_{-2\text{V}} \cong \frac{\Delta v_{DS}}{\Delta i_D}$$

$$= \frac{20 - 10}{(4.1 - 3.9)10^{-3}} = 50\ \text{k}\Omega$$

Assuming $V_p \cong -6$ V and $I_{DSS} \cong 10$ mA, use Eq. 17-17 to calculate the value of g_m at $V_{GS} = -2$ V and compare to the value obtained graphically.

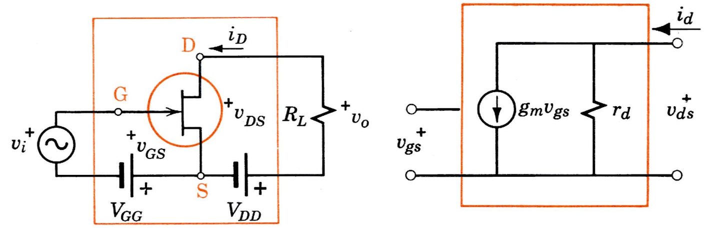
$$g_m = -\frac{2I_{DSS}}{V_p}\left(1 - \frac{V_{GS}}{V_p}\right)$$

$$= -\frac{2 \times 10^{-2}}{-6}\left(1 - \frac{-2}{-6}\right) = 2200\ \mu\mho$$

approximately the same as the measured value.

Small-Signal Model

On the diagram of the elementary amplifier in Fig. 17.7a, v_{GS}, v_{DS}, and i_D are total instantaneous quantities representing small ac variations superimposed on steady or dc values. The dc values fix the quiescent point and determine the values of the FET parameters; the ac values correspond to the signals, which are of primary interest. We wish to devise a small-signal model that will hold only for ac values and that will enable us to predict the performance of the FET in an ac application such as this elementary amplifier.



(a) Elementary amplifier (b) Small-signal model

Figure 17.7 Derivation of a small-signal model of an FET.

For differential changes, Eq. 17-13 states that

$$di_D = \frac{\partial i_D}{\partial v_{GS}} dv_{GS} + \frac{\partial i_D}{\partial v_{DS}} dv_{DS} \tag{17-19a}$$

Replacing the differentials by small signals and using the FET parameters, Eq. 17-19a becomes

$$i_d = g_m v_{gs} + \frac{1}{r_d} v_{ds} \tag{17-19b}$$

This relationship holds equally well for effective values of sinusoidal quantities (or to the corresponding phasors) and Eq. 17-19b becomes

$$I_d = g_m V_{gs} + \frac{1}{r_d} V_{ds} \tag{17-19c}$$

This relation must be satisfied by the ac circuit if it is to be a functional model. Equation 17-19b says that the drain signal current consists of two parts. In the corresponding parallel circuit, one branch contains a controlled current source $g_m v_{gs}$ directly proportional to the input signal voltage, and the other branch carries a current v_{ds}/r_d directly proportional to output voltage. The circuit of Fig. 17.7b is the logical result.

Example 3

Predict the voltage gain of the amplifier of Fig. 17.7a if $R_L = 5\text{ k}\Omega$.

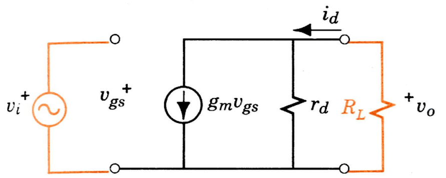


Figure 17.8 Voltage gain calculation.

The amplifier is replaced by the circuit model of Fig. 17.8 where $g_m = 2300\,\mu\mho$ and $r_d = 50\text{ k}\Omega$.

Since $r_d \gg R_L$,

$$v_o = -g_m v_i (R_L \| r_d) \cong -g_m v_i R_L$$

and the voltage gain is

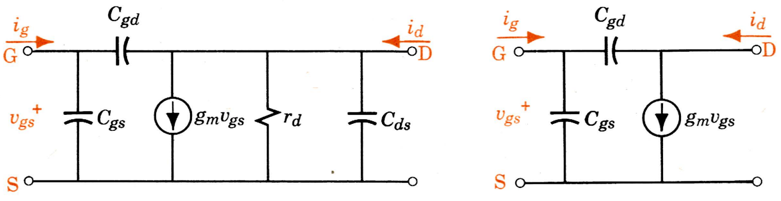
$$\frac{v_o}{v_i} = -g_m R_L = -2300 \times 10^{-6} \times 5 \times 10^3 \cong -11$$

The minus sign indicates an inversion or phase reversal of the output signal.

The use of models in the analysis of FET amplifiers is illustrated in Example 3. One conclusion is that for practical values of R_L (limited by available supply voltage V_{DD}), voltage gains are small; much larger gains can be obtained with BJT amplifiers. (See Example 4.) Note that with an increase in $|v_{GS}|$, g_m increases and r_d decreases. Note also that r_d typically is so large compared to practical values of R_L that reasonably good predictions can be made with a model consisting of controlled-current source $g_m v_{gs}$ alone.

High-Frequency Model

At high frequencies consideration must be given to the charging currents associated with various capacitive effects. In Fig. 17.9a, the small-signal model has been modified to take into account capacitances inherent in field-effect transistors. C_{gs} and C_{gd} represent the capacitances between gate and source and gate and drain; in a JFET these arise from the reverse-biased *pn* junctions and will be of the order of 1 to 10 pF. C_{ds} represents the drain to source capacitance and will be small, say 0.1 to 1 pF. As predicted by this model, the voltage gain of an FET amplifier drops off as frequency increases. A quantitative analysis of high-frequency performance is given in Chapter 18; the simplified model of Fig. 17.9b is usually satisfactory.



(a) Complete

(b) Simplified

Figure 17.9 High-frequency FET models.

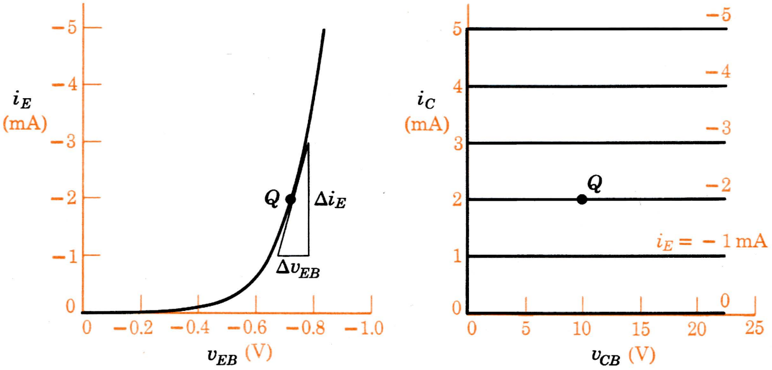


Figure 17.10 Idealized transistor characteristics (CB).

Junction Transistors

Figure 13.17 shows the graphical analysis of an elementary large-signal amplifier. For small signals, a linear circuit model is much more convenient in both analysis and design. As with the FET, our objective is a simple model that will predict the performance of a properly biased transistor (BJT) insofar as small ac variations are concerned.

An Idealized Model (Common-Base)

The *i-v* characteristics of a typical transistor are displayed in Fig. 13.12. The general relations can be expressed in terms of voltages or currents as the independent variables. We have the choice of whether to deal with impedances, admittances, or a hybrid combination in defining parameters. If we choose the last, the general relations for a common-base transistor are

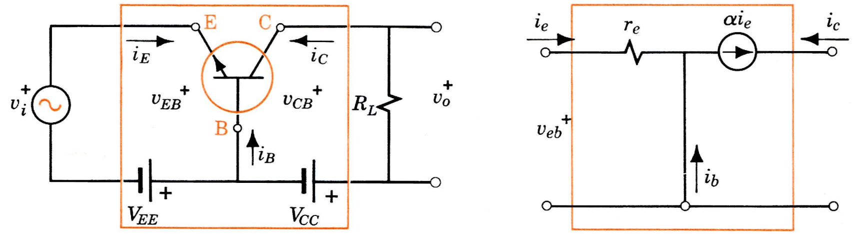
$$\begin{cases} v_{EB} = f_1(i_E, v_{CB}) \\ i_C = f_2(i_E, v_{CB}) \end{cases} \tag{17-20}$$

Before deriving a general model, let us reason out a circuit model from the idealized *v-i* characteristics of Fig. 17.10. The emitter characteristics are those of a theoretical diode and the collector characteristics indicate a constant α and negligible I_{CBO}. For operation at a quiescent point Q defined by $I_E = -2$ mA and $V_{CB} = 10$ V, a small change Δv_{EB} produces a small change Δi_E that, in turn, produces a small change Δi_C (for nomenclature, see Table 13-1). The small-signal behavior of this idealized transistor is defined by

$$\begin{cases} v_{eb} = r_e i_e \\ i_c = -\alpha i_e \end{cases} \tag{17-21}$$

where $$r_e = \frac{\Delta v_{EB}}{\Delta i_E} = \frac{v_{eb}}{i_e} = \text{the dynamic junction resistance} \tag{17-22}$$

$$-\alpha = \frac{\Delta i_C}{\Delta i_E} = \frac{i_c}{i_e} = \text{the forward current-transfer ratio} \tag{17-23}$$



(a) Elementary CB amplifier (b) Small–signal model

Figure 17.11 Derivation of a small-signal model of a transistor (CB).

Since $i_b = -(i_e + i_c)$ by Kirchhoff's current law, the circuit representation of these relations is shown in Fig. 17.11b. The emitter-base junction of a transistor behaves like a junction diode and r_e is given by Eq. 17-8. At room temperature,

$$r_e = \frac{0.025}{I_E} = \frac{25}{I_E(\text{in mA})}\text{ohms} \tag{17-24}$$

This gives a simple accurate means for estimating r_e in practical transistors.

Example 4

(a) A transistor with $\alpha = 0.98$ is to operate at room temperature with $I_E = 0.5$ mA. Estimate the dynamic emitter resistance.

By Eq. 17-24

$$r_e = \frac{25}{I_E(\text{in mA})} = \frac{25}{0.5} = 50\ \Omega$$

(b) If operation increases the junction temperature by 60°K, predict the new emitter resistance.

For this 20% increase in junction temperature T,

$$kT/e \cong 25(300 + 60)/300 = 30\text{ mV}$$

and

$$r_e' = 30/0.5 = 60\ \Omega$$

(c) The transistor is to operate at room temperature in the amplifier circuit of Fig. 17.11a where $R_L = 5\text{ k}\Omega$. For an output signal $v_o = 2$ V, determine the signal input current i_i and the voltage gain.

For calculations involving small signals, the transistor is replaced by the model of Fig. 17.11b.

In Fig. 17.12, $i_e = i_i$ and $v_o = \alpha i_e R_L = \alpha i_i R_L$. The necessary signal current is

$$i_i = \frac{v_o}{\alpha R_L} = \frac{2}{0.98 \times 5000} = 410\ \mu\text{A}$$

The input voltage required is

$$v_i = r_e i_i = 50 \times 410 \times 10^{-6} = 0.0205\text{ V}$$

The voltage gain is

$$\frac{v_o}{v_i} = \frac{2}{0.0205} = 98$$

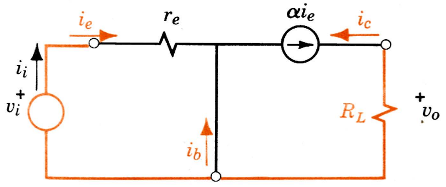


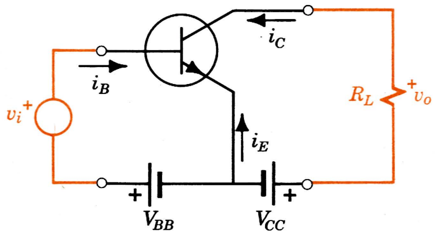
Figure 17.12 Application of CB model.

The Common-Emitter Configuration

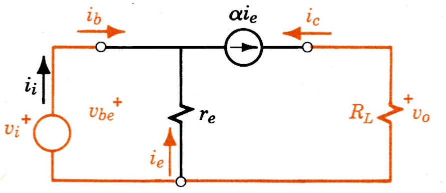
One advantage of a model is that it may provide new insight into the behavior of the actual device. Insofar as small signals are concerned, the linear model of Fig. 17.12 can replace the actual physical transistor and its dc supplies. If the transistor is reconnected as shown in Fig. 17.13a, the base-emitter terminals become the input port and the collector-emitter terminals become the output port. To represent this common-emitter configuration, the model is merely rearranged, but the device performance is dramatically changed by this rearrangement.

Example 5

The transistor of Example 4 is reconnected as a common-emitter amplifier. Determine the signal input current i_i now required.



(a) Schematic diagram



(b) Circuit model

Figure 17.13 Small-signal circuit model reconnected in the CE configuration.

As shown in Fig. 17.13,

$$i_e = -i_b + \alpha i_e \quad \text{or} \quad i_e = \frac{-i_b}{1-\alpha}$$

Then

$$v_o = \alpha i_e R_L = -\frac{\alpha}{1-\alpha} i_b R_L$$

For the same 5-kΩ load resistance and the same 2-V output, the input signal is

$$i_i = i_b = -\frac{(1-\alpha)v_o}{\alpha R_L} = -\frac{0.02 \times 2}{0.98 \times 5000} = -8.2\,\mu\text{A}$$

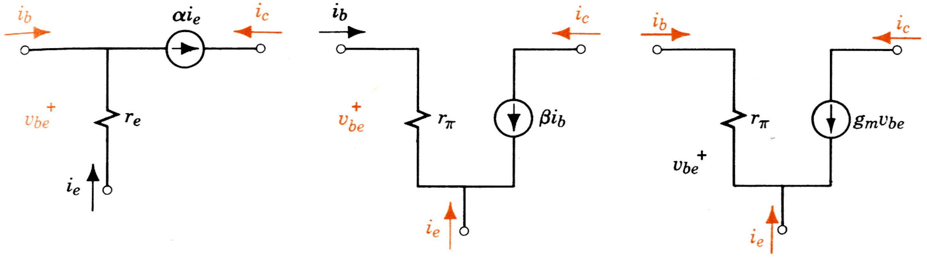
In comparison to the 410 μA required in the common-base amplifier, only 8.2 μA is required here.

There is an effective current gain of

$$\frac{410}{-8.2} \cong -50$$

The negative sign indicates a 180° phase reversal of the output signal with respect to the input.

The possibility of a current gain as well as a voltage gain is an important advantage of the common-emitter amplifier. Because this configuration is used so extensively, we are justified in deriving a special model. Although the model of Fig. 17.14a could be used, it is not convenient because the controlled source is not a function of the input current; also, the character of the current gain is obscured in this model. A better form is that of Fig. 17.14b, where β and r_π are to be determined. As a first step, we note that $i_b + i_e = \alpha i_e$. Solving,



(a) Generator controlled by emitter current

(b) Generator controlled by base current

(c) Generator controlled by base–emitter voltage

Figure 17.14 Circuit models for an idealized transistor (CE).

$$i_e = -\frac{i_b}{1-\alpha} \quad \text{or} \quad \alpha i_e = -\frac{\alpha}{1-\alpha} i_b = -\beta i_b \tag{17-25}$$

where β is now the small-signal current gain (CE). For most purposes, $\beta_{\text{ac}} \cong \beta_{\text{dc}}$.

The current generator in Fig. 17.14b is controlled by the input current i_b. Now the character of the current gain is more clear. The effect of a variation in base current (a signal input to the base) appears in the collector circuit as a controlled source whose magnitude is $\alpha/(1-\alpha)$ times as great. For a typical transistor with $\alpha = 0.98$, $\beta = 0.98/0.02 = 49$.

The new relation between the input voltage and current is obtained by noting that

$$v_{be} = -v_{eb} = -r_e i_e = +r_e \frac{i_b}{1-\alpha} \tag{17-26}$$

The input resistance for the common-emitter configuration is therefore,

$$r_\pi = \frac{v_{be}}{i_b} = \frac{r_e}{1-\alpha} = (\beta+1) r_e \cong \beta r_e \tag{17-27}$$

Since $I_C = \alpha I_E \cong I_E$, Eq. 17-24 can be used to obtain

$$r_\pi \cong \beta \frac{25}{I_C(\text{in mA})} \text{ohms} \tag{17-28}$$

The "r_π-β model" of Fig. 17.14b represents the small-signal behavior of an idealized transistor in the common-emitter configuration. (It could have been deduced directly from an idealization of the curves of Fig. 13.15 where $I_B \cong I_C/\beta$.) Alternatively, the output current can be defined in terms of an input voltage (Fig. 17.14c). Since $v_{be} = r_\pi i_b = \beta r_e i_b$,

$$\beta i_b = \frac{1}{r_e} v_{be} = g_m v_{be} \tag{17-29}$$

where g_m, the small-signal transconductance for the BJT, is given by

$$g_m = \frac{1}{r_e} = 40 I_C \tag{17-30}$$

Under certain conditions the common-base arrangement is superior, but where current amplification as well as voltage amplification is important, the common-emitter amplifier possesses obvious advantages. (See Example 6.) Under other circumstances (discussed in Chapter 18), the common-collector configuration is advantageous.

Example 6

An amplifier employing a transistor with $\beta = 50$ and operating at $I_C = 0.5$ mA in the common-emitter configuration is to deliver a signal current of 0.2 mA rms to a load resistance of 5 kΩ. Estimate the signal current gain and voltage gain.

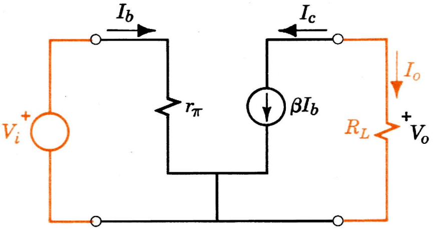


Figure 17.15 Application of the CE transistor model.

For this estimate, the r_π-β model of Fig. 17.15 is satisfactory. Since $I_o = -I_c = -\beta I_b$, the signal current gain is

$$\frac{I_o}{I_b} = -\beta = -50$$

The necessary input signal current (rms) is

$$I_b = \frac{-I_o}{\beta} = -\frac{0.2 \times 10^{-3}}{50} = -4 \times 10^{-6} = -4\,\mu\text{A}$$

By Eq. 17-28 the input resistance is

$$r_\pi = \beta r_e = \beta \frac{25}{I_C} = 50 \times \frac{25}{0.5} = 2500\,\Omega$$

$$\therefore\ V_i = I_b r_\pi = -4 \times 10^{-6} \times 2.5 \times 10^3 = -10\text{ mV}$$

The voltage gain is

$$\frac{V_o}{V_i} = \frac{+I_o R_L}{V_i} = -\frac{2 \times 10^{-4} \times 5000}{10 \times 10^{-3}} = -100$$

The Hybrid-π Model

For more precise predictions of transistor behavior, more sophisticated models are necessary. In Fig. 17.16, the *base-spreading resistance* r_b takes into account the ohmic resistance of the long, thin base region. Typically, r_b varies from 50 to 150 Ω; lacking other information, a value of 100 Ω can be assumed with little error if r_π is relatively high. The resistance r_μ accounts for a very small current due to changes in collector-base voltage; this effect is negligibly small in most cases and r_μ will be ignored in our work. The resistance r_o provides for the small change in i_C with v_{CE} for given values of i_B (see Fig. 13.15).

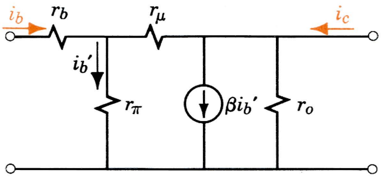
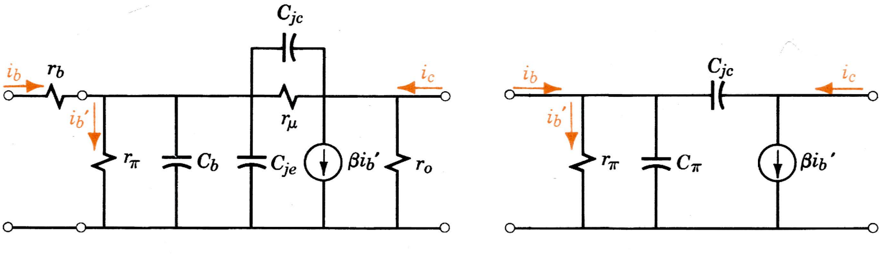


Figure 17.16 The more precise hybrid-π circuit model.

In this model, proposed by L. J. Giacoletto, the circuit elements are arranged in a π-configuration known as the *hybrid-π* model.† It is particularly useful, when properly modified, in predicting the high-frequency behavior of transistors. Junction capacitances C_{je} and C_{jc} and capacitance C_b accounting for base-charging current can be connected in parallel with r_π and r_μ as shown in the high-frequency hybrid-π model of Fig. 17.17a. If r_b and r_o are neglected and r_μ ignored, and if C_b and C_{je} are combined into C_π, the simplified high-frequency model of Fig. 17.17b results. This model will be discussed in more detail in connection with amplifier performance in Chapter 18. (Note that the r_π-β model is a simple hybrid-π for moderate frequencies.)



(a) Complete

(b) Simplified

Figure 17.17 High-frequency hybrid-π models.

The General h-Parameter Model

For any configuration and any small-signal application, a general linear model can be used to represent the transistor and predict its performance. Hybrid parameters are most commonly used and are most frequently supplied by manufacturers. The chief advantages of h-parameters are the ease with which they can be determined in the laboratory and the facility with which they can be handled in circuit calculations.

Choosing input current i_I and output voltage v_O as the independent variables and rewriting Eq. 17-11 in conventional transistor notation,

---

† It is customary to use the *internal* base-emitter voltage $v_{be}' = i_b' r_\pi$ as the control variable here.

$$\begin{cases} v_I = f_1(i_I, v_O) \\ i_O = f_2(i_I, v_O) \end{cases} \tag{17-31}$$

For the common-emitter configuration, for example, v_I is the base-emitter voltage and i_O is the collector current.

Because of the nonlinearity of these functions, no simple model can be formulated to predict total voltages and currents. Since we are interested in small-signal behavior, let us consider differential changes in the variables. Taking the total differential, these become

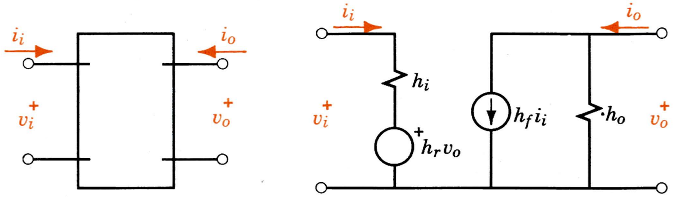
$$\begin{cases} dv_I = \dfrac{\partial v_I}{\partial i_I} di_I + \dfrac{\partial v_I}{\partial v_O} dv_O \\ \\ di_O = \dfrac{\partial i_O}{\partial i_I} di_I + \dfrac{\partial i_O}{\partial v_O} dv_O \end{cases} \tag{17-32}$$

Equation 17-32a says that a change in input current di_I and a change in output voltage dv_O both contribute to a change in input voltage dv_I. The effectiveness of the partial contributions is dependent on the coefficients $\partial v_I/\partial i_I$ and $\partial v_I/\partial v_O$; these *partial derivatives* of input voltage with respect to input current and output voltage and the corresponding partial derivatives of output current are the h-parameters of a small-signal circuit model.

Noting that the differential quantities di_I and dv_O correspond to small signals i_i and v_o, Eqs. 17-32 can be written as

$$\begin{cases} v_i = \dfrac{\partial v_I}{\partial i_I} i_i + \dfrac{\partial v_I}{\partial v_O} v_o = h_i i_i + h_r v_o \\ \\ i_o = \dfrac{\partial i_O}{\partial i_I} i_i + \dfrac{\partial i_O}{\partial v_O} v_o = h_f i_i + h_o v_o \end{cases} \tag{17-33}$$

where h_i = input impedance with output short-circuited (ohms),
h_f = forward transfer current ratio with output short-circuited,
h_r = reverse transfer voltage ratio with input open-circuited,
h_o = output admittance with input open-circuited (mhos).



(a) Two-port network (b) Circuit model

Figure 17.18 General h-parameter representation.

The first of Eqs. 17-33 says that the input voltage is the sum of two components. This indicates a series combination of an impedance drop $h_i i_i$ and a controlled voltage source $h_r v_o$ directly proportional to the output voltage. The second equation indicates a parallel combination of a controlled current source $h_f i_i$ and an admittance current $h_o v_o$. The h-parameter model of Fig. 17.18b follows from this line of reasoning.

The general model is applicable to any transistor configuration. In a specific case, the parameters are identified by a second subscript b, e, or c, depending on whether the base, emitter, or collector is the common element.

Example 7

The transistor of Example 6 has the following h-parameters: $h_{ie} = 2500\ \Omega$, $h_{re} = 4 \times 10^{-4}$, $h_{fe} = 50$, and $h_{oe} = 10\ \mu℧$ as shown in Fig. 17.19.

For $R_L = 5\ \text{k}\Omega$ and $I_o = 0.2$ mA rms, estimate the current gain and voltage gain of the amplifier and compare with the results of Example 4.

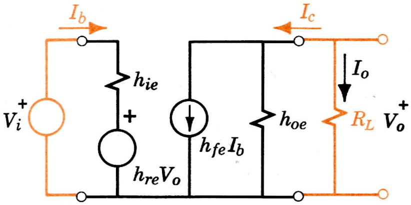


Figure 17.19 Application of general h-parameters in the CE configuration.

Applying the current-divider principle,

$$I_o = -I_c = -\frac{1/R_L}{h_{oe} + 1/R_L} h_{fe} I_b$$

The current gain is

$$\frac{I_o}{I_b} = -\frac{h_{fe}/R_L}{h_{oe} + 1/R_L} = -\frac{50/5000}{(10 + 200)10^{-6}} = -47.6$$

The required input signal is

$$I_b = I_o/(-47.6) = 200/(-47.6) = -4.2\ \mu\text{A}$$

In the input loop,

$$\begin{aligned} V_i &= I_b h_{ie} + h_{re} V_o \\ &= -4.2 \times 10^{-6} \times 2500 + 4 \times 10^{-4} \times 1 \\ &= (-10.5 + 0.4)\ 10^{-3} = -10.1\ \text{mV} \end{aligned}$$

Since $V_o = I_o R_L = 0.2 \times 10^{-3} \times 5 \times 10^3 = 1$ V,

$$\frac{V_o}{V_i} = \frac{1}{-0.0101} = -99$$

The simpler model introduces only small errors.

Determination of h-Parameters

The h-parameters are small-signal values implying small ac variations about operating points defined by dc values of voltage and current. They could be determined from families of characteristic curves. In actual practice, however, the parameters are determined experimentally by measuring the ac voltages or currents that result from ac signals introduced at the appropriate locations. For example, to determine h_{fe}, a small ac voltage is applied between base and emitter and the resulting ac currents in the base lead and the collector lead are measured (the collector circuit must be effectively short-circuited to ac). Then h_{fe} is the ratio of collector current to base current and is just equal to β. Similar measurements permit the determination of the other small-signal parameters. Note that h_{ie} is $r_\pi + r_b \cong r_\pi$ and $h_{ib} \cong r_e$.

Since transistor parameters vary widely with operating point, measurements are made under standard conditions. Usually, values are for a frequency of 1000 Hz at room temperature (25°C) with an emitter current of 1 mA and a collector-base voltage of 5 V. Transistor manufacturers publish typical, minimum, and maximum values of the parameters along with average static characteristics. Typical values are shown in Table 17-1. An engineer designing a transistor amplifier, for example, must arrange the circuit so that the performance of the amplifier is within specifications despite expected variations in transistor parameters. (In Chapter 18 we see how feedback can be used to stabilize performance.)

Table 17-1 Typical Transistor Parameters

				2N699B		
	2N525	2N1613	2N3114	Minimum	Typical	Maximum
$h_{ie}(\mathrm{k}\Omega)$	1.4	2.2	1.5		2.8	
$h_{re}(\times 10^{-4})$	3.37	3.6	1.5		3.5	
h_{fe}	44	55	50	35	70	100
$h_{oe}(\mu\mho)$	27	12.5	5.3		11	
$h_{ib}(\Omega)$	31	27	27	20	27	30
$h_{rb}(\times 10^{-4})$	5	0.7	0.25		0.5	1.25
h_{fb}	−0.978	−0.98	−0.981		−0.987	
$h_{ob}(\mu\mho)$	0.60	0.16	0.09	0.1	0.12	0.5

Comparison of BJTs with FETs

The much higher values of g_m permit much higher voltage gains with the BJT as compared to the FET. On the other hand, the much higher values of input resistance for the FET permit lower "loading" of the source as compared to the BJT. The more linear transfer characteristic of the BJT permits larger signal swings without distortion; the square-law transfer characteristic of the FET is advantageous in modulation and mixing. In integrated circuit form, the MOSFET is smaller and simpler and therefore cheaper than the BJT.

A Communication System

If part of the output of an amplifier is fed back into the input in just the right way (discussed in Chapter 18), it is possible to obtain self-sustained oscillations. Such an amplifier is called an *oscillator* or a *signal generator*. With the addition of the oscillator to our repertory of electronic devices we have all the essential elements of a telemetry or communication system (Fig. 17.20).

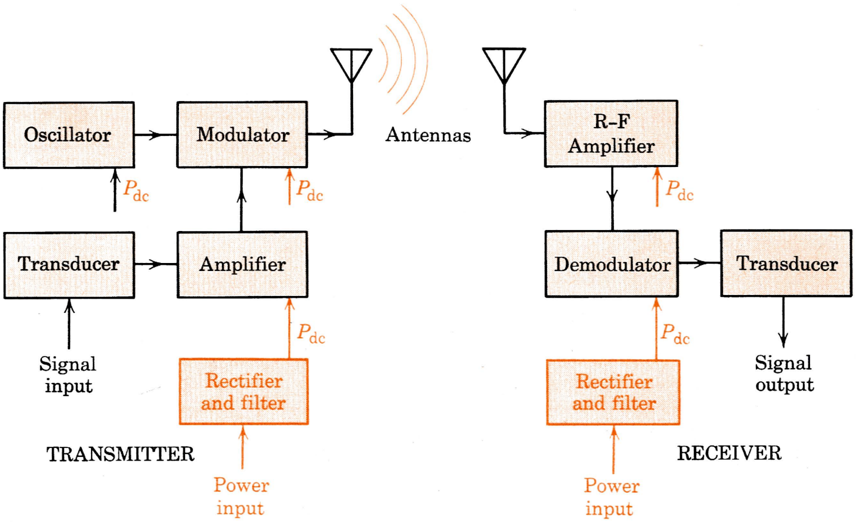

Oscillator
Modulator
Antennas
R-F Amplifier
P_{dc}
P_{dc}
P_{dc}
Transducer
Amplifier
Demodulator
Transducer
P_{dc}
P_{dc}
Signal input
Rectifier and filter
Rectifier and filter
Signal output
TRANSMITTER
RECEIVER
Power input
Power input

Figure 17.20 Elements of a communication system.

In a radio broadcasting system the signal input is an audible variation in air pressure on a microphone or a mechanical vibration of a phonograph needle. The transducer, a microphone or pickup, converts the acoustic vibration into an electrical signal that, when amplified, is available for modulating. The radio-frequency carrier signal is generated in the oscillator and combined with the audiofrequency signal in the modulator. Usually the modulator stage also performs amplification so that a high-level amplitude-modulated signal is fed into the transmitting antenna. Energy is radiated from the antenna in the form of an electromagnetic wave, part of which is intercepted by the receiving antenna. The selected carrier and sidebands are amplified and fed into the demodulator. The detector output, usually amplified further, is converted by the transducer (a loudspeaker) into a replica of the original input signal. The power required may be obtained from 60-Hz ac lines, converted into dc power by rectifier-filter combinations, and supplied to each stage of the communication system.

SUMMARY

- The function performed by a diode depends on bias voltage and signal amplitude.
  A "small signal" is one to which a given device responds in a linear fashion.

- A modulator superimposes information on a carrier by modifying the carrier amplitude, frequency, or phase.
  An AM wave consisting of carrier and sidebands is described by

$$v_{\mathrm{AM}} = V_C \sin \omega_C t - \frac{mV_C}{2} \cos(\omega_C + \omega_M)t + \frac{mV_C}{2} \cos(\omega_C - \omega_M)t$$

  A demodulator recovers the original signal from the modulated wave.

- The small-signal performance of complicated and highly nonlinear electronic devices can be predicted using relatively simple linear models.
  For small signals superimposed on steady values, the significant diode parameter is the dynamic resistance $r_{\mathrm{ac}} = dv/di$, the reciprocal of the slope of the i-v characteristic at the operating point.

- To derive the small-signal circuit model of a nonlinear two-port:

  1. Write the governing equation in terms of the chosen variables.
  2. Take the total differential to obtain the equations for small variations.
  3. Note that differentials can be approximated by small signals, and write the corresponding equations for small signals.
  4. Interpret the equations in terms of an electrical circuit with appropriate parameters, including passive elements and controlled sources.

- For a field-effect transistor, the input current is negligible and the small-signal model consists of transconductance g_m and drain resistance r_d (Fig. 17.21). For a junction transistor, the general model is complicated; at moderate frequencies the hybrid-parameter CE model is as shown in Fig. 17.22.
  In many transistor applications, a satisfactory model consists simply of an input resistance r_π and a controlled current source βI_b (Fig. 17.23).

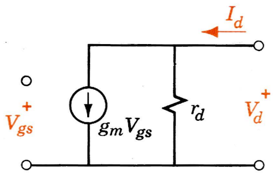


Figure 17.21

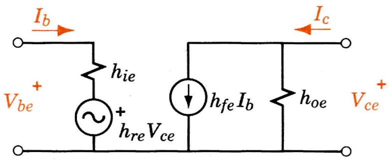


Figure 17.22

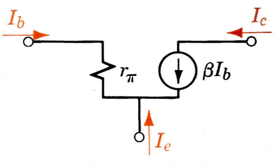


Figure 17.23

- Because individual transistors vary widely from the norm, it is customary to use parameter value ranges published by the manufacturer, minimizing the effects of parameter variations by proper circuit design.

REVIEW QUESTIONS

1. Explain the differences between V/I, $\Delta V/\Delta I$, dV/dI, and $\partial V/\partial I$.
2. Explain the difference between static resistance and dynamic resistance.
3. Sketch and label the spectrum of frequencies transmitted by your favorite AM station during a typical music program.
4. Explain how a nonlinear device produces amplitude modulation. Demodulation.
5. Why is a high frequency needed for a radio broadcast carrier?
6. How many telephone conversations (two-way) can be carried on a telephone cable that transmits frequencies up to 8 MHz? How many television programs?
7. Why is the ac model of an FET simpler than an equally precise model of a BJT?
8. In analytic geometry, what does $z = f(x, y)$ represent in general? Could $i_b = f(v_c, v_b)$ be similarly represented? What is the graphical interpretation of $\partial z/\partial y$?
9. Sketch a set of FET characteristics and define, graphically, the two parameters.
10. What are the criteria for selecting the quiescent point in an amplifier?
11. Does the circuit of Fig. 17.7b contain any information not in Eq. 17-19b? Why use the circuit?
12. What is meant by: "The results of a graphical analysis of performance from characteristic curves lack generality"?
13. Why are no dc supplies such as batteries shown in the linear models? What is the ac impedance of a battery?
14. Explain how a current gain is obtained in a transistor amplifier.
15. Define the h-parameters graphically in terms of common-base characteristic curves. Why are h-parameters not calculated this way?

EXERCISES

1. A 900-kHz carrier has its amplitude varied sinusoidally from zero to $2V_C$ at a rate of 3000 times per second. Sketch the resulting modulated wave and list the frequencies of the components appearing in the transmitted wave.
2. A signal is described by the equation

   $$v = (10 + 5 \sin 1000t + 2 \sin 2000t) \sin 10^6 t$$

   List the frequencies (rad/s) present in the signal and show them on a frequency spectrum. Identify the carrier and the upper and lower sidebands.
3. An amplitude-modulated wave is described by the equation

   $$v = 20 \cos 1.98 \times 10^6 t + 40 \sin 2.00 \times 10^6 t - 20 \cos 2.02 \times 10^6 t$$

   Identify the carrier frequency (rad/s) and the frequencies of any modulating signals.
4. Channel 4, a typical TV channel, extends from 66 to 72 MHz. A ruby laser operates in the red region with a wavelength of 7500 Å. If the useful bandwidth of a laser beam is equal to 0.2% of its carrier frequency, how many TV channels could be carried on a single laser beam?
5. Two signals $a = \sin \omega t$ and $b = 5 \sin 20\omega t$ are multiplied together.
   (a) Show on a frequency spectrum all the components of the product.
   (b) If a voltage equal to the product is fed into a device for which $i_o = v + 0.2v^2$, show on a frequency spectrum all the components of the output current.
6. A signal $v = A \sin 100{,}000t + B \sin 103{,}000t$ is applied to the input of an FET whose output is defined by $i_o = av^2$.
   (a) List the frequencies of the components appearing in the output.

   (b) Show the components on a frequency spectrum.
   (c) Under what conditions does this represent demodulation, i.e., the separation of an audible signal from a carrier?
   (d) Sketch a filter circuit to pass only the audible component to load R_L.

7. Derive a general expression for the dynamic resistance of a device for which $i = v + 0.01\,v^2$ A and evaluate r_{ac} at $V = 10$ V. Predict $i(t)$ for $v(t) = 10 + \sin \omega t$ V.

8. A silicon diode is characterized by $I_s = 10$ nA. Predict the dc forward resistance and the dynamic resistance at:
   (a) $T = 25°\text{C}$ and forward current $= 0.1$ mA.
   (b) $T = 100°\text{C}$ and forward current $= 0.05$ mA.

9. Starting from the second pair of Eqs. 17-10, and assuming a linear two-port network, derive a circuit model with admittance parameters (Y_{11}, etc.).

10. Starting from the second pair of Eqs. 17-10, and assuming a nonlinear two-port device, derive a small-signal model with admittance parameters (y_{11}, etc.).

11. Represent the MOSFET of Fig. 13.4c by a small-signal model and estimate the parameters at a point defined by $v_{GS} = 0$ and $v_{DS} = 8$ V.

12. For a JFET with $g_{mo} = 4000\,\mu\mho$ and $V_p = -4$ V, predict g_m at:
   (a) $V_{GS} = -1$ V; (b) $I_D = 2$ mA.

13. For a DE MOSFET with $I_{DSS} = 10$ mA and $g_{mo} = 5000\,\mu\mho$, predict g_m at:
   (a) $I_D = 5$ mA; (b) $V_{GS} = -1$ V.

14. The MOSFET of Exercise 13 ($r_d = 50\text{ k}\Omega$) is used in the circuit of Fig. 17.7a with $R_L = 5\text{ k}\Omega$ and $V_{GG} = -2$ V.
   (a) Specify I_D and V_{DD} for operation at $V_{GS} = -2$ V and $V_{DS} = 8$ V.
   (b) Predict the voltage gain.

15. For the MOSFET of Exercise 11 specify the value of load resistance to provide a voltage gain of 10.

16. The JFET of Exercise 12 is used in the circuit of Fig. 17.7a with $V_{GG} = -1$ V, $V_{DD} = 30$ V, and $R_L = 4\text{ k}\Omega$. Predict the quiescent drain current I_D and the input voltage V_i(rms) to produce an output voltage $V_o = 30$ mV.

17. A semiconductor two-port is described by the equations

$$\begin{cases} i_1 = 10^{-6}\, e^{40v_1} \\ i_2 = -0.98 i_1 + 10^{-6} v_2. \end{cases}$$

Devise a small-signal model for this device at an operating point defined by $i_1 = 2$ mA and $v_2 = -5$ V.

18. A two-port electronic device is defined by the following equations:

$$\begin{cases} i_1 = 0.1 v_1 + 0.1\, dv_1/dt \\ v_2 = 5 v_1 + 20 i_2 \end{cases}$$

where currents are in amperes and voltages in volts. Devise a suitable circuit model and label the circuit elements with the appropriate values.

19. A transistor with an α of 0.99 is used in the circuit of Fig. 17.11.
   (a) Is this an *npn* or *pnp* transistor?
   (b) Is it biased properly for normal operation? If not, make the appropriate changes in battery polarity.
   (c) Redraw the circuit using a simple small-signal model for an idealized transistor to replace the portion within the colored lines.
   (d) Derive a general expression for voltage gain $A_V = v_o/v_i$.
   (e) Evaluate A_V for operation at $I_E = 0.5$ mA and $R_L = 2\text{ k}\Omega$.

20. The transistor of Exercise 19 is reconnected in the circuit of Fig. 17.13a. Use the same model and repeat parts (c), (d), and (e).

21. Compare the voltage gains, current gains, and power gains (i.e., the product of voltage gain times current gain) for the amplifiers of Exercises 19 and 20 and draw a conclusion.

22. A transistor with $\beta = 100$ is used in an elementary common-emitter amplifier with $I_C = 2$ mA and $R_L = 5\text{ k}\Omega$.
   (a) Use the r_π-β model to predict the signal voltage V_s required for an output voltage $V_o = 2$ V rms.
   (b) Repeat part (a) assuming the signal source has an internal resistance of 1 kΩ in series with V_s.

23. A transistor has the following parameters known to ±10% accuracy: $r_b = 100\,\Omega$, $r_\pi = 2500\,\Omega$, $r_\mu = 20\text{ M}\Omega$, $\beta = 100$, and

$r_o = 200$ kΩ. Predict the voltage gain of the elementary amplifier of Fig. 17.13 with $R_L = 2$ kΩ:
   (a) Using the circuit model of Fig. 17.16.
   (b) Using the circuit model of Fig. 17.14.
   (c) Compare the results of parts (a) and (b) and draw a conclusion.

24. A 2N3114 transistor operating at 25°C and $I_E = 1$ mA is used in an elementary amplifier to provide 2 V rms across $R_L = 5$ kΩ. The input source can be represented by V_s in series with $R_s = 1$ kΩ.
   (a) Using the general h-parameter model, specify V_s.
   (b) Using the r_π-β model, repeat part (a) and draw a conclusion.

25. For certain parameters, the manufacturer of the 2N699B specifies guaranteed minimum and maximum values (Table 17-1). Determine the corresponding range of values of r_π. Is high precision justifiable in calculations using these parameters?

26. A 2N1613 is to be used ($I_C = 1$ mA, $V_{CB} = 5$ V) in a common-emitter amplifier to provide 3 V rms output across a load resistance R_L. The available open-circuit signal is 20 mV from a source with an internal resistance of 500 Ω.
   (a) Draw and label the circuit with an h-parameter model.
   (b) Simplify the model where appropriate to permit an approximate analysis.
   (c) Specify the necessary load resistance and the collector supply voltage V_{CC}.

27. Compare the voltage gain possibilities of an FET rated at $I_{DSS} = 4$ mA and $g_{mo} = 5000\ \mu$℧ and a BJT rated at $\beta = 100$ at $V_{CE} = 5$ V, both operating at $I_D = I_C = 1$ mA. Predict the actual voltage gains with $R_C = R_D = 5$ kΩ assuming the voltage source has an output resistance of 2500 Ω, and draw a conclusion.

PROBLEMS

1. When a nonlinear device is used as a detector, the important factor is the second-degree term in Eq. 17-1. Prove that an FET whose characteristic is defined by $i = av^2$ in combination with a suitable filter will recover the modulating signal from an amplitude-modulated wave.
2. The diode-capacitor circuit of Fig. 12.12 can be used as a detector for AM signals. Assume that the input voltage v is the modulated carrier of Fig. 17.2 and that time constant R_LC is large compared to the period T_C but small compared to T_M.
   (a) Sketch v_L, the output signal.
   (b) If the "load," an earphone, includes an inductive reactance that is large at the carrier frequency, sketch the output current i_L.
3. A *mixer* is widely used in radio and TV receivers to obtain a signal with a frequency equal to the *difference* in the frequencies of two input signals. Because of its square-law transfer characteristics an FET makes an ideal mixer. Demonstrate that if the sum of two signals at frequencies ω_x and ω_y is applied to the gate-source terminals of a DE MOSFET, the output current will contain a component at frequency $\omega_x - \omega_y$. What are the other components of the output and how could they be eliminated?
4. A practical JFET amplifier circuit is shown in Fig. 17.24.

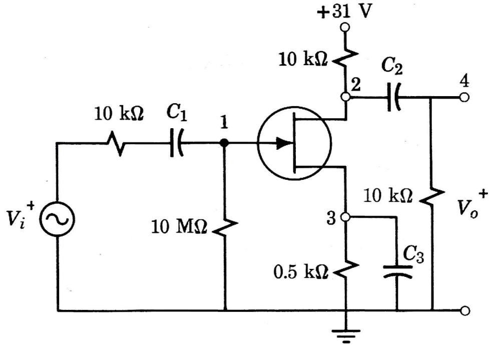


Figure 17.24

   (a) Redraw the circuit for dc analysis, assume $V_{DS} = 10$ V and estimate the bias voltage V_{GS} and quiescent current I_D.
   (b) Redraw the circuit for ac analysis and

evaluate the parameters given that $g_{mo} = 8\ \text{m}\mho$ and $I_{DSS} = 8\ \text{mA}$.

(c) Predict the voltage gain.

(d) Draw one cycle of a sinusoidal $v_i(\omega t)$ and, on the same graph, show the voltages with respect to ground at points 1, 2, 3, and 4.

5. In the circuit of Fig. 17.7a, the input source consists of V_i in series with $R_i = 10\ \text{k}\Omega$ and $R_L = 5\ \text{k}\Omega$. The JFET can be represented by the model of Fig. 17.19 with $g_m = 5\ \text{m}\mho$, $r_d = 100\ \text{k}\Omega$, $C_{gs} = C_{gd} = 10\ \text{pF}$ and $C_{ds} = 1\ \text{pF}$.

(a) Draw a small-signal model of the amplifier and evaluate the capacitive reactances at $f = 10\ \text{kHz}$.

(b) Redraw the circuit model neglecting any capacitances that have a small effect compared to the associated resistances and calculate the voltage gain V_o/V_i.

(c) Recalculate the reactances at $f = 1\ \text{MHz}$ and redraw the amplifier circuit as in part (b). Assume $V_o = 50\ \text{mV}$ (rms) and calculate the necessary voltage V_{gs} across C_{gs} and the voltage across C_{gd}. Calculate the currents through C_{gs} and C_{ds} and the necessary voltage V_i (rms).

(d) Calculate the voltage gain at 1 MHz and draw a conclusion.

6. A transistor amplifier is terminated in a load resistance R_L. Using the general h-parameter circuit model (Fig. 17.18b), derive an expression for voltage gain v_o/v_i.

7. A transistor amplifier is terminated in a load resistance R_L. Using the general h-parameter circuit model (Fig. 17.18b), derive an expression for input resistance v_i/i_i.

8. An elementary transistor amplifier is to be used as a pre-amplifier for a "hi-fi" set. The output of the phonograph pickup (low internal impedance) is only 2 mV, but 0.4 V is required as the input to the power amplifier that drives the loudspeaker. "Design" the pre-amplifier by specifying a suitable transistor and collector current and the appropriate load resistor and supply voltage.

9. In a particular transistor application (Fig. 17.25), the usual load resistance is omitted and an emitter resistance R_E is inserted. V_{BB} and V_{CC} are adjusted to bring operation to an appropriate quiescent point. Replace the transistor with an r_π-β model and derive expressions for: (a) v_E in terms of v_i (voltage gain), and (b) input resistance $r_i = v_i/i_b$. Assuming $\beta = 50$, $r_\pi = 1200\ \Omega$, and $R_E = 2000\ \Omega$, evaluate r_i and compare with the input resistance of an ordinary common-emitter amplifier (Fig. 17.13).

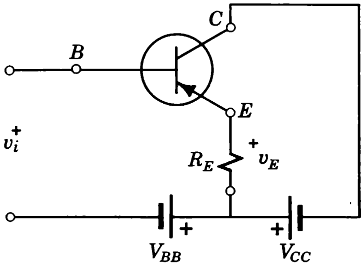


Figure 17.25

Chapter 18

Small-Signal Amplifiers

RC Coupled Amplifiers
Frequency Response
Multistage Amplifiers
Feedback Amplifiers

To amplify the signal from a strain gage or a microphone to a power level adequate to drive a recorder or a loudspeaker usually requires several stages of amplification. If the final amplifier is a power transistor, appreciable driving current is needed and several stages of current amplification may be required. The low-level voltage or current amplification is provided by "small-signal" amplifiers that respond linearly.

In a "small-signal" amplifier, the input signals are small in comparison to the dc bias and the resulting output swings are small in comparison to the quiescent operating values of voltage and current. Under small-signal operation, bias is not critical and amplitude distortion is easily avoided. Although graphical analysis is possible, the linear models developed in Chapter 17 are much more convenient to use.

We have already used the linear models to predict the performance of single-stage amplifiers at moderate frequencies. Now we are going to investigate the gain of untuned amplifiers as a function of frequency and consider the effect of cascading several stages. Field-effect and bipolar transistors with *RC* coupling are commonly employed for small-signal amplification in the frequency range from a few cycles per second to several megahertz, and therefore they are emphasized in this discussion. Although BJTs and FETs operate on different physical bases, their external behavior is quite similar and the same approach is used in predicting their frequency responses.

RC Coupled Amplifiers

Each stage of a multistage amplifier consists of an electronic two-port, a biasing network, and the coupling circuits. Commonly the output of a transistor is coupled to the next stage by means of a load resistor and a coupling capacitor.

Frequency-Response Curve

The voltage gain of one stage of an *RC*-coupled amplifier is shown in Fig. 18.1. The gain is relatively constant over the *midfrequency range,* but falls off at lower frequencies and at higher frequencies. In ordinary amplifiers, the frequency response curve is symmetric if frequency is plotted on a logarithmic scale as shown. The *bandwidth* is defined by lower and upper *cutoff frequencies* or *half-power frequencies* f_1 and f_2, just as in the frequency response of a resonant circuit. At the half-power frequencies, the response is 70.7% of the midfrequency gain, or $A_1 = A_2 = A_O/\sqrt{2}$, and the phase angle $\theta = \pm 45°$.

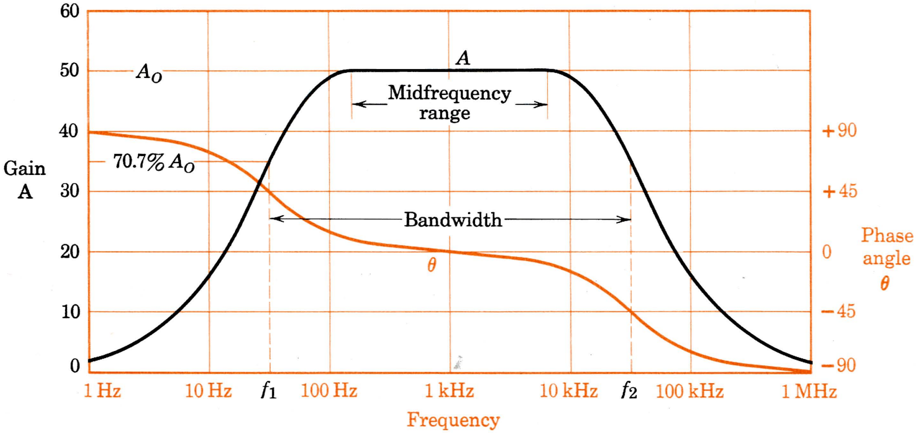
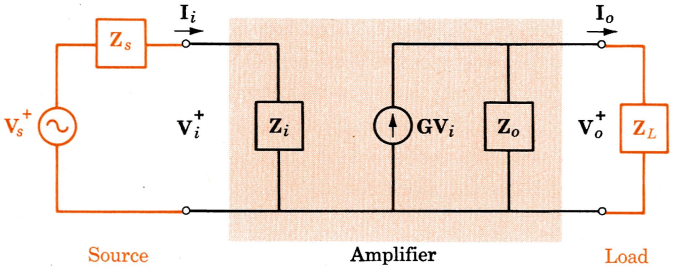


Figure 18.1 Frequency response of an *RC*-coupled single-stage amplifier.

A Generalized Amplifier

In the general amplifier circuit of Fig. 18.2, the input signals come from a "source," maybe a phono pickup or a previous amplifier stage, that is charac-

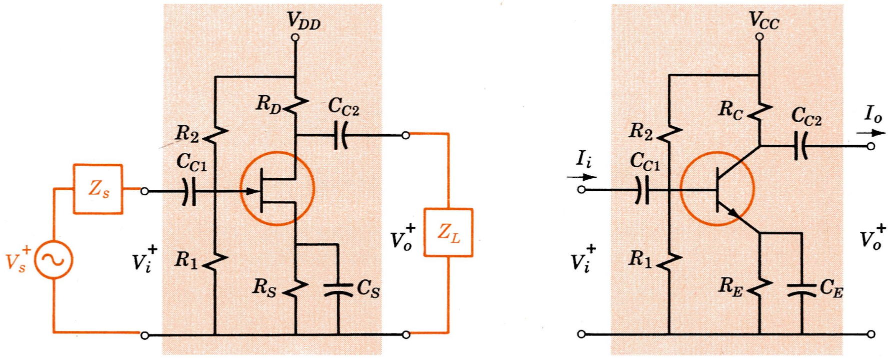


$\mathbf{I}_i$
$\mathbf{I}_o$
$\mathbf{Z}_s$
$\mathbf{V}_s$
$\mathbf{V}_i$
$\mathbf{Z}_i$
$\mathbf{GV}_i$
$\mathbf{Z}_o$
$\mathbf{V}_o$
$\mathbf{Z}_L$
Source
Amplifier
Load

Figure 18.2 A generalized one-stage amplifier circuit.

terized by an open-circuit voltage $\mathbf{V}_s$ and Thévenin equivalent impedance $\mathbf{Z}_s$. The input signals consist of sinusoids of various amplitudes and frequencies so variables are shown as effective-value phasors. The "one-stage" amplifier is characterized by input and output impedances $\mathbf{Z}_i$ and $\mathbf{Z}_o$ and a controlled source (controlled by input current or voltage) with gain $\mathbf{A}$ where $\mathbf{Z}_i$, $\mathbf{Z}_o$, and $\mathbf{A}$ are complex functions of frequency. The "load" of the amplifier may be a transducer or a following amplifier stage. Given the components of the input signal and the characteristics of the source and load, we can predict the output if we know the frequency response of the amplifier.

Typical small-signal amplifier circuits are shown in Fig. 18.3; these are "*RC*-coupled amplifiers" because of coupling capacitors C_{C1} and C_{C2} and their associated resistances. Our approach is to replace these complicated circuits by linear models that respond in the same way to ac signals. We assume that the transistors are operating at proper quiescent points and the bias voltages and currents are of no concern. Bias resistors R_S and R_E are assumed to be properly bypassed so they do not appear in the models of Fig. 18.4. Since batteries are short circuits to ac signals, they are omitted. On the other hand, unintentional elements such as junction capacitances and wiring capacitances (C_W), which do not appear on the wiring diagram, must be shown in the amplifier models. Reflecting the most common situation, the source impedance is simply R_s and the load is represented by C_L and R_L in parallel.

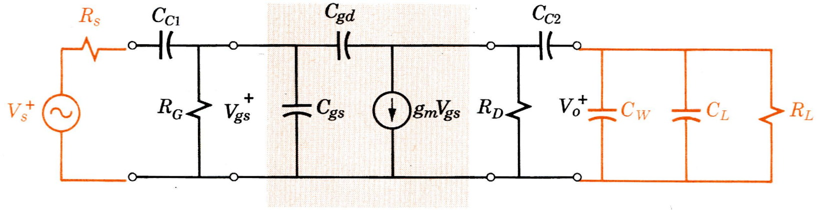


(a) FET (b) BJT

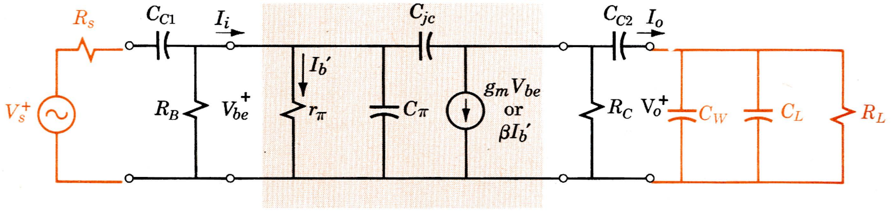
Figure 18.3 Single-stage *RC*-coupled amplifiers.

Small-Signal Amplifier Models

The device models are those derived in Chapter 17. (See Figs. 17-9b and 17-17b.) For the FET, C_{gs} and C_{gd} are gate-source and gate-drain capacitances and g_m is the transductance; r_d has been omitted since usually $r_d \gg R_D$. For



(a) FET



(b) BJT

Figure 18.4 Linear circuit models of *RC*-coupled amplifiers.

the BJT, C_π is the effective input capacitance representing junction and diffusion capacitances, r_π is the small-signal input resistance, and C_{jc} is the collector-base junction capacitance. Output resistance r_o has been omitted since $r_o \gg R_C$. Base spreading resistance r_b has been neglected on the assumption that $r_b \ll R_s + r_\pi$; therefore, $V_{be}' = V_{be}$. The current source can be considered to be controlled by the base current or the base-emitter voltage since $g_m V_{be} = \beta I_b'$ where g_m is the transconductance and β is the small-signal current gain.

The determination of the values of individual elements in Fig. 18.4 is quite difficult; the specific values depend upon such factors as quiescent voltage and current, operating temperature, characteristics of the following stage, and configuration of the physical circuit. The important thing from our viewpoint here is that an actual amplifier can be replaced, insofar as small signals are concerned, by a linear circuit model that can be analyzed by methods we have already mastered. Fortunately, the models for FET and BJT are similar in form and the same analysis can be used for both.

These linear models hold fairly well for frequencies from a few cycles per second to several megahertz. It would be possible to perform a general analysis on the circuits as they stand, but it is much more convenient, and more instructive, to consider some simplifying approximations. We recall that for series and parallel combinations of resistance R and capacitance C,

$$Z_{\text{ser}} = \sqrt{R^2 + (1/\omega C)^2} = R\sqrt{1 + (1/\omega CR)^2} \tag{18-1}$$

$$Z_{\text{par}} = \frac{1}{\sqrt{(1/R)^2 + (\omega C)^2}} = \frac{R}{\sqrt{1 + (\omega CR)^2}} \tag{18-2}$$

We note that in Eq. 18-1 if $\omega CR \geq 10$, $Z_{\text{ser}} \cong R$ within 0.5%. Also, if in Eq. 18-2 $\omega CR \leq 0.1$, $Z_{\text{par}} \cong R$.

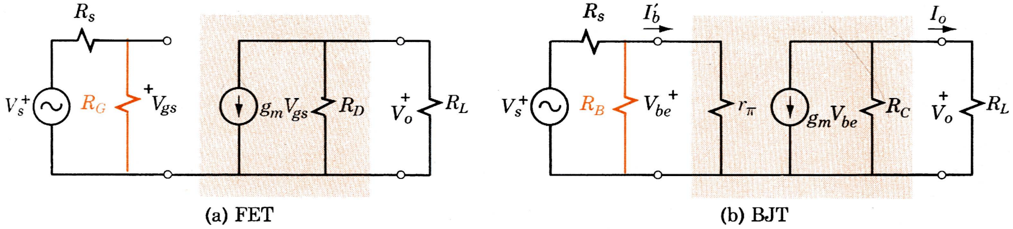
The fact that gain is independent of frequency over the midfrequency range (Fig. 18.1) indicates that there is a range of frequencies over which the capacitive effects are negligible. This is the case in practical amplifiers (see Exercise 1), and it leads to another important conclusion. Since ωC increases with frequency, series capacitances become important only at frequencies lower than the midfrequencies and parallel capacitances become important only at frequencies higher than the midfrequencies.

Frequency Response

The analysis of series and parallel RC combinations leads to the convenient conclusion that the frequency response of an RC amplifier can be divided into three regions. In the midfrequency range, the capacitances can be neglected. In the low-frequency range, the series capacitances C_C must be considered. In the high-frequency range, the parallel capacitances must be considered.

Midfrequency Gain

In a properly designed untuned amplifier, there is a range of frequencies over which the general models of Fig. 18.4 can be replaced by the purely resistive circuits of Fig. 18.5.



(a) FET (b) BJT

Figure 18.5 Mid-frequency models of amplifiers.

FET. Since $R_G \gg R_s$, R_G can be omitted in most cases. Where the parallel combination of R_D and R_L is represented by R_o, the output voltage is

$$V_o = -g_m V_{gs}(R_D \| R_L) = -g_m V_s R_o \tag{18-3}$$

and the midfrequency voltage gain is

$$A_{VO} = \frac{V_o}{V_s} = -g_m R_o \tag{18-4}$$

BJT. For proper stability R_B should be small compared to βR_E (Eq. 16-12) and for negligible loading of the input circuit R_B should be large compared to r_π. Here we assume that the ideal condition $r_\pi \ll R_B \ll \beta R_E$ is achieved so the circuit is stable *and* simple to analyze. (See Exercise 4.) Where the parallel combination of R_C and R_L is represented by R_o, the output voltage is

$$V_o = -g_m V_{be}(R_C \| R_L) = -g_m \frac{r_\pi}{r_\pi + R_s} V_s R_o = -\beta \frac{V_s}{r_\pi + R_s} R_o \qquad (18\text{-}5)$$

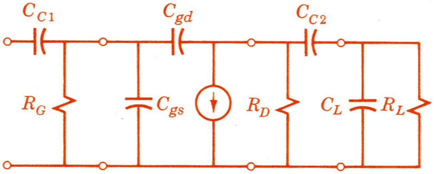
and the midfrequency voltage gain is

$$A_{VO} = \frac{V_o}{V_s} = -\frac{r_\pi}{r_\pi + R_s} g_m R_o = -\beta \frac{R_o}{r_\pi + R_s} \qquad (18\text{-}6)$$

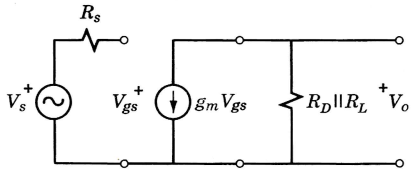
Performance of these one-stage amplifiers is influenced by the characteristics of the source (R_s) and the load (R_L). In the multistage amplifiers discussed later, the source and the load for each stage are other similar stages.

Example 1

A JFET with $g_m = 2000\ \mu\mho$, $C_{gs} = 20$ pF, and $C_{gd} = 2$ pF is used in an RC-coupled amplifier with $R_G = 1\ \text{M}\Omega$, $R_S = 2\ \text{k}\Omega$, and $R_D = 5\ \text{k}\Omega$. Signal source resistance $R_s = 1\ \text{k}\Omega$ and load resistance $R_L = 5\ \text{k}\Omega$. Coupling capacitance $C_{C1} = C_{C2} = 1\ \mu\text{F}$ and $C_S = 20\ \mu\text{F}$; the load capacitance (including wiring) $C_L = 20$ pF. Predict the voltage gain V_o/V_s at $\omega = 10^4$ rad/s ($f = 1590$ Hz).



(a) General model



(b) Midfrequency model

Figure 18.6 Mid-frequency simplification.

At the given frequency,

$$1/\omega C_S = 1/10^4 \times 20 \times 10^{-6} = 5\ \Omega$$

Therefore, R_S is effectively shorted and the model is as shown in Fig. 18.6a. Since C_{C1} and C_{C2} are effectively in series with R_G and R_L, respectively, the reactances

$$1/\omega C_C = 1/10^4 \times 10^{-6} = 100\ \Omega$$

are negligibly small (Eq. 18-1). With C_{C1} and C_{C2} omitted, C_L is in parallel with $R_o = R_D \| R_L = 2500\ \Omega$ and C_{gs} is in series with R_s. The reactances

$$1/\omega C_L = 1/\omega C_{gs} = 1/10^4 \times 20 \times 10^{-12} = 5\ \text{M}\Omega$$

are so large that C_L and C_{gs} can be omitted. The reactance

$$1/\omega C_{gd} = 1/10^4 \times 2 \times 10^{-12} = 50\ \text{M}\Omega$$

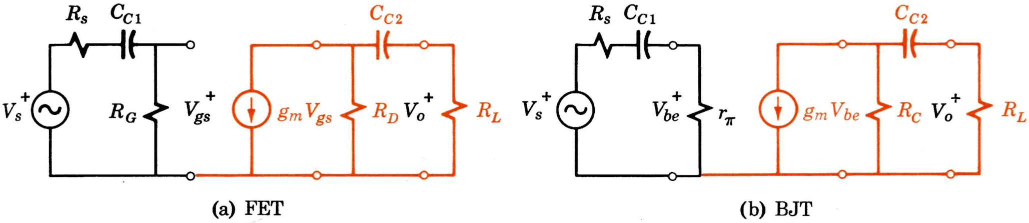
is so large that a negligible current will flow in C_{gd}. In other words, 1590 Hz is in the "midfrequency" range of this amplifier and the gain (Eq. 18-4) is

$$A_{VO} = -g_m R_o = -2000 \times 10^{-6} \times 2500 = -5.$$

Low-Frequency Response

Below the midfrequencies, the susceptances of the parallel capacitances are negligibly small, but the reactances of the coupling capacitances C_C become increasingly important.

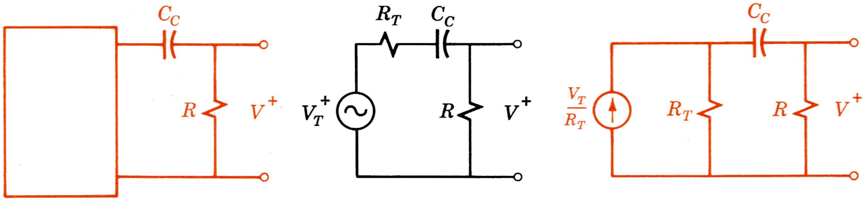
Coupling Capacitors. The general models can be reduced to the simpler low-frequency models shown in Fig. 18.7, where R_G has been retained because it is significant but R_B has been omitted since it is large compared to r_π. With this simplification, the FET and BJT models are analogous and one analysis will serve for both.



(a) FET (b) BJT

Figure 18.7 Low-frequency models of amplifiers.

The input and output circuits are "high-pass filters" (see Fig. 7.10) of slightly different form. The general circuit (Fig. 18.8a) can be represented by the Thévenin form as in the input circuit or by the Norton form as in the output circuit; the two forms are equivalent. For midfrequencies where C_C is effectively shorted, the output of the Thévenin circuit is



(a) General circuit (b) Thévenin form (c) Norton form

Figure 18.8 High-pass filter circuits.

$$V_O = \frac{R}{R + R_T} V_T \tag{18-7}$$

At low frequencies where C_C is significant,

$$\mathbf{V}_L = \frac{R}{R + R_T - j\dfrac{1}{\omega C_C}}\mathbf{V}_T = \frac{\dfrac{R}{R + R_T}\mathbf{V}_T}{1 - j\dfrac{1}{\omega C_C(R + R_T)}} = \frac{\mathbf{V}_O}{1 - j\dfrac{1}{\omega C_C(R + R_T)}} \tag{18-8}$$

after dividing through by $R + R_T$. In words, the low-frequency output of either filter circuit is related to the midfrequency output by a complex factor dependent on frequency and an RC product. As frequency is decreased, a larger fraction of voltage V_T appears across C_C and the voltage V at the output is reduced. The *cutoff* or *half-power frequency* where $|V_L/V_O| = 1/\sqrt{2}$ is defined by

$$\omega_{\text{co}} = \frac{1}{(R + R_T)C_C} \tag{18-9}$$

The behavior of the FET or BJT amplifiers in Fig. 18.7 at low frequencies can now be predicted. At a frequency defined by

$$\omega_{11} = \frac{1}{C_{C1}(R_s + R_G)} \quad \text{or} \quad \omega_{11} = \frac{1}{C_{C1}(R_s + r_\pi)} \tag{18-10}$$

the input voltage V_{gs} or V_{be} will be down to 70% of V_s and the possible amplifier gain will be correspondingly reduced. At a frequency defined by

$$\omega_{12} = \frac{1}{C_{C2}(R_D + R_L)} \quad \text{or} \quad \omega_{12} = \frac{1}{C_{C2}(R_C + R_L)} \tag{18-11}$$

the output voltage V_o will be down to 70% of $g_m V_{gs} R_o$ or $g_m V_{be} R_o$ and the amplifier gain will be correspondingly reduced. The overall low-frequency voltage gain for the FET amplifier can be expressed as

$$\mathbf{A}_L = \mathbf{A}_O \cdot \frac{1}{1 - j\dfrac{1}{\omega C_{C1}(R_s + R_G)}} \cdot \frac{1}{1 - j\dfrac{1}{\omega C_{C2}(R_D + R_L)}} \tag{18-12}$$

or, for either FET or BJT, the *relative gain* at low frequencies is

$$\frac{\mathbf{A}_L}{\mathbf{A}_O} = \frac{1}{1 - j\omega_{11}/\omega} \cdot \frac{1}{1 - j\omega_{12}/\omega} \tag{18-13}$$

To predict the behavior of a given circuit, determine ω_{11} and ω_{12}; the higher of the two determines the cutoff frequency (unless they are very close together; see Exercise 10). To design a circuit for a specified behavior, calculate C_{C1} and C_{C2} from Eqs. 18-10 and 11, specify the larger at its calculated value, and make the smaller several times its calculated value.

Bypass Capacitors. The foregoing analysis of low-frequency response assumes that C_S (or C_E) effectively bypasses R_S (or R_E) at the lowest frequency of interest. In practice, it turns out that C_S must be much larger than C_{C1} and C_{C2} and, in most cases, C_S is the important element in determining the lower cutoff frequency. The practical approach is to assume that the coupling capacitors are still effective at the frequencies where C_S becomes critical. With this assumption, the low-frequency FET amplifier model is as shown in Fig. 18.9.

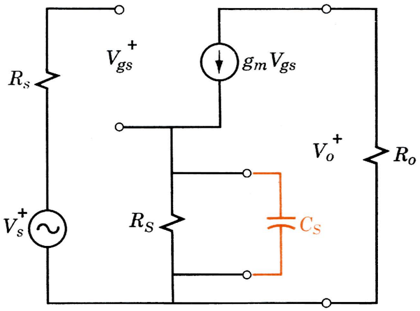


Figure 18.9 Low-frequency amplifier model where C_S is critical.

At low frequencies, the voltage gain is reduced because the current $g_m\,V_{gs}$ flowing through Z_S (the parallel combination of R_S and C_S) develops a voltage that subtracts from the signal voltage V_s. Summing voltages around the input loop, $\mathbf{V}_s - \mathbf{V}_{gs} - g_m\,\mathbf{V}_{gs}\,\mathbf{Z}_S = 0$ or the gate voltage is reduced to

$$\mathbf{V}_{gs} = \frac{\mathbf{V}_s}{1 + g_m\,\mathbf{Z}_S} \tag{18-14}$$

and the voltage gain V_o/V_s is reduced by the factor $1 + g_m\,\mathbf{Z}_S$. The lower cutoff frequency occurs when $1 + g_m\,\mathbf{Z}_S$ has a magnitude of $\sqrt{2}$. Since $\mathbf{Z}_S$ is a complex quantity, there is no simple expression for the proper value of C_S. Instead we shall use a useful technique to determine an approximate expression.†

With R_S unbypassed, the gain is low (but not zero, because Z_S has a maximum value of R_S). With R_S perfectly bypassed, the gain is $A_V = g_m R_o$. Therefore, let us specify C_S to be large enough so that at f_1 the reactance $1/\omega_1 C_S$ is equal to the *effective resistance* bypassed. We define the effective resistance as the Thévenin equivalent resistance looking in at the terminals of R_S across which C_S is connected. Letting $Z_S = R_S$ in Eq. 18-14,

$$V_{OC} = g_m V_{gs} R_S = g_m \frac{V_s}{1 + g_m R_S} R_S = \frac{g_m R_S}{1 + g_m R_S} V_s \tag{18-15}$$

Since $I_{SC} = g_m V_s$ with R_S shorted, the Thévenin equivalent is

$$R_T = \frac{V_{OC}}{I_{SC}} = \frac{R_S}{1 + g_m R_S} = \frac{1}{g_m + 1/R_S} = R_S \| 1/g_m \tag{18-16}$$

or the effective resistance is the parallel combination of R_S and $1/g_m$. Therefore the design criterion is

$$\omega_1 = 2\pi f_1 = \frac{1}{C_S(R_S \| 1/g_m)} = \frac{g_m + 1/R_S}{C_S} \tag{18-17}$$

---

† Suggested by Malvino; a general expression is given on p. 449 of Schilling and Belove. See References at the end of Chapter 16.

Example 2

Specify C_S for a lower cutoff frequency of 20 Hz in the circuit of Fig. 18.9 where $g_m = 2000\ \mu\mho$ and $R_S = 2\ \text{k}\Omega$. Show that the voltage gain at 20 Hz is down to approximately 70% of its midfrequency value.

By Eq. 18-17,

$$C_S = \frac{g_m + 1/R_S}{2\pi f_1} = \frac{0.002 + 0.\ 005}{2\pi \times 20} \cong 20\ \mu\text{F}$$

For this value of C_S and $f = 20$ Hz,

$$g_m \mathbf{Z}_S = \frac{g_m}{\dfrac{1}{R_s} + j\omega C} = \frac{0.002}{0.0005 + j0.0025}$$

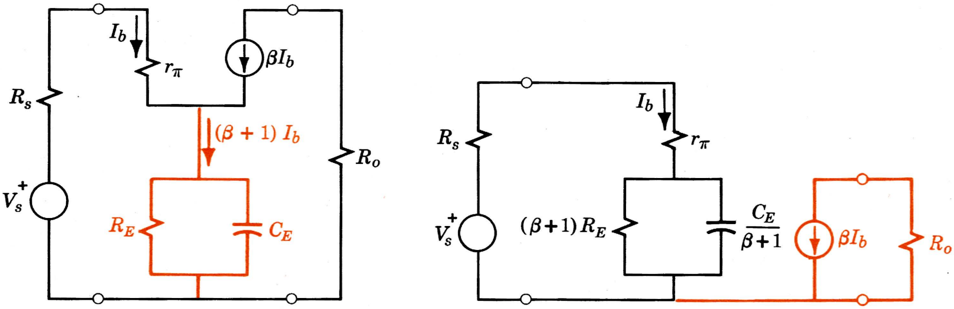
$$= 0.784 \underline{/-78.7^\circ}$$

$$1 + g_m \mathbf{Z}_S = 1 + (0.15 - j0.77) = 1.15 - j0.77$$

$$= 1.38 \underline{/33^\circ}$$

The magnitude of $1 + g_m \mathbf{Z}_S$ is $1.38 \cong \sqrt{2} = 1.41$; therefore f_1 is approximately the lower cutoff frequency and Eq. 18-17 is justified.

In the BJT amplifier, C_E is the important element in determining the lower cutoff frequency. Here again an exact analysis is complicated and an approximate expression is satisfactory for our purposes. The approach employed in establishing the design criterion for C_S in the FET amplifier would work, but using a different approach may be more enlightening. As before, we assume that the coupling capacitors are effective at the frequency where C_E becomes critical and draw the low-frequency amplifier model as shown in Fig. 18.10a.



(a) (b)

Figure 18.10 The low-frequency model where C_E is critical.

The current $(\beta + 1)\mathbf{I}_b$ flowing through $\mathbf{Z}_E$ (the parallel combination of R_E and C_E) develops a voltage that, in effect, opposes the signal voltage $\mathbf{V}_s$. Summing the voltages around the left-hand loop and solving for the current,

$$\mathbf{I}_b = \frac{\mathbf{V}_s}{R_s + r_\pi + (\beta + 1)\mathbf{Z}_E} \tag{18-18a}$$

where

$$(\beta + 1)\mathbf{Z}_E = \frac{\beta + 1}{\dfrac{1}{R_E} + j\omega C_E} = \frac{1}{\dfrac{1}{(\beta + 1)R_E} + j\omega\dfrac{C_E}{\beta + 1}} \tag{18-18b}$$

Equations 18-18a and b are satisfied by the circuit model of Fig. 18-10b. If we neglect the current through the relatively large resistance $(\beta + 1)R_E$,

$$\mathbf{I}_b = \frac{\mathbf{V}_s}{R_s + r_\pi - j\,\dfrac{\beta + 1}{\omega C_E}} \tag{18-19}$$

If the degrading effect of $\mathbf{Z}_E$ is to be small, the reactance of $C_E/(\beta + 1)$ must

Example 3

In the circuit of Fig. 18.11, $R_s = 2\text{ k}\Omega$, $R_E = 1\text{ k}\Omega$, and $C_E = 50\ \mu\text{F}$. R_1 and R_2 are very large. For $r_\pi = 1.5\text{ k}\Omega$ and $\beta = 50$, estimate the lower cutoff frequency and check the validity of the assumption made in deriving Eq. 18-20. For $R_C = 5\text{ k}\Omega$ and $R_L = 10\text{ k}\Omega$, specify appropriate values for C_{C1} and C_{C2}.

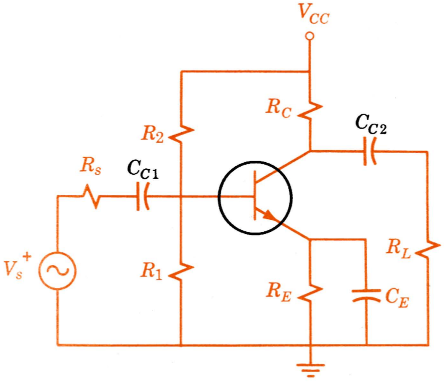


Figure 18.11 Low-frequency response.

By Eq. 18-21 (see Fig. 18.10),

$$f_1 = \frac{\omega_1}{2\pi} = \frac{\beta + 1}{2\pi C_E(R_s + r_\pi)}$$

$$= \frac{50 + 1}{2\pi \times 5 \times 10^{-5}(2 + 1.5)10^3} = 46\text{ Hz}$$

The admittance of $(\beta + 1)R_E$ is

$$\frac{1}{(\beta + 1)R_E} = \frac{1}{51 \times 10^3} \cong 2 \times 10^{-5}\,\mho$$

which when squared (see Eq. 18-2) is negligibly small in comparison to the square of

$$\frac{\omega_1 C_E}{\beta + 1} = \frac{1}{R_s + r_\pi} = \frac{1}{3.5 \times 10^3} = 29 \times 10^{-5}\,\mho$$

For the same value of f_1, Eqs. 18-10 and 11 indicate

$$C'_{C1} = \frac{1}{2\pi f_1(R_s + r_\pi)} = \frac{1}{2\pi \times 46(3500)} \cong 1\ \mu\text{F}$$

therefore, specify a value of, say, $5C'_1 = 5\ \mu\text{F}$

$$C'_{C2} = \frac{1}{2\pi f_1(R_C + R_L)} = \frac{1}{2\pi \times 46(15{,}000)} \cong 0.23\ \mu\text{F}$$

therefore, specify a value of, say, $5C'_{C2} = 1\ \mu\text{F}$.

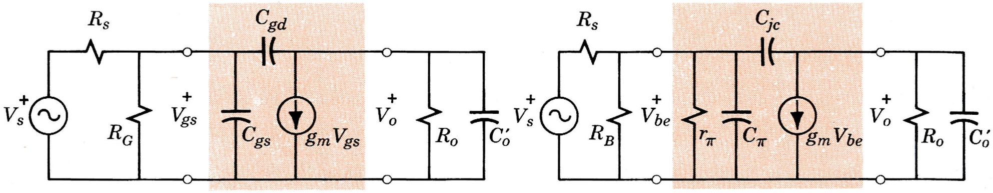
be small compared to the resistances with which it is in series. Current I_b will drop to 70.7% of its midfrequency value and the gain will drop to 70% of its midfrequency value at frequency ω_1 where

$$\frac{1}{\omega_1 C_E/(\beta + 1)} = R_s + r_\pi \tag{18-20}$$

therefore, the design criterion is

$$\omega_1 = 2\pi f_1 = \frac{\beta + 1}{C_E(R_s + r_\pi)} \tag{18-21}$$

In practice, C_E is specified to provide the desired low-frequency response and then C_{C1} and C_{C2} are selected to be several times the values indicated by Eqs. 18-10 and 11. Note that the cutoff frequency is just a convenient measure of the low-frequency response of the amplifier and does not imply that no amplification occurs below this frequency. (See Example 3 on preceding page.)



(a) FET (b) BJT

Figure 18.12 High-frequency models of amplifiers.

High-Frequency Response

Above the midfrequencies, the reactances of coupling and bypass capacitances are negligibly small, but the susceptances of the other capacitances become increasingly important. Under these conditions, the high-frequency versions of the amplifier circuit models are as shown in Fig. 18.12, where $R_o = R_D \| R_L$ or $R_C \| R_L$ and $C_o' = C_W + C_L$. The analysis of these analogous circuits is complicated by the presence of C_{gd} and C_{jc} that provide negative feedback from the output to the input and therefore tend to reduce the gain at high frequencies.

As a first step, let us simplify the circuit using a technique first suggested by J. M. Miller. In Fig. 18.13, the voltage

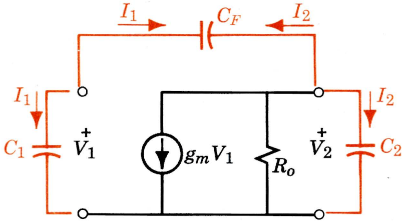


Figure 18.13 Miller capacitances.

across feedback capacitance C_F causing current I_1 is

$$V_1 - V_2 = V_1 - (-g_m R_o)V_1 = V_1(1 + g_m R_o) = V_1(1 + A) \quad (18\text{-}22)$$

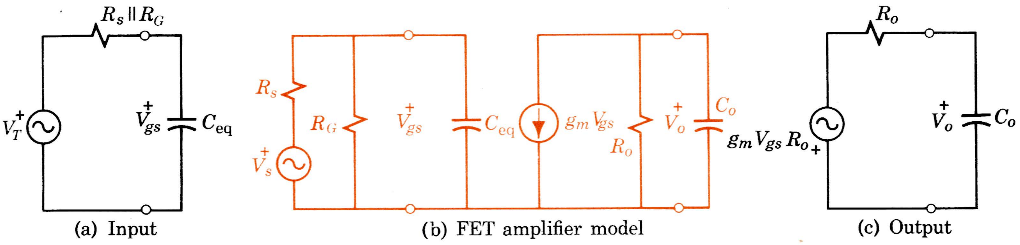
where $V_2/V_1 = -A = -g_m R_o$ is the gain of this circuit. This current could be accounted for in the input circuit by a capacitance

$$C_1 = C_F(1 + A) = C_F(1 + g_m R_o) \quad (18\text{-}23)$$

under a voltage V_1. Following the same reasoning, the current I_2 could be accounted for by a capacitance

$$C_2 = C_F(1 + 1/A) = C_F(1 + 1/g_m R_o) \cong C_F \quad (18\text{-}24)$$

under a voltage V_2 since $V_2 - V_1 = V_2(1 + 1/A) \cong V_2$. When these substitutions are made and the "Miller capacitances" are combined with C_{gs} and C_o', respectively, the simplified circuit model is as shown in Fig. 18.14b where $C_{\text{eq}} = C_{gs} + C_{gd}(1 + A)$ and $C_o = C_{gd} + C_W + C_L$.



(a) Input (b) FET amplifier model (c) Output

Figure 18.14 Low-pass filter circuits.

FET. In the Thévenin equivalent (Fig. 18.14a), $R_T = R_s \| R_G \cong R_s$ since $R_G \gg R_s$, in general. Now the input circuit has the form of a *low-pass filter.* (The parallel RC circuit is the dual of the series RL circuit of Fig. 7.9.) At midfrequencies, $V_{gs} = V_T \cong V_s$. At high frequencies where C_{eq} is significant,

$$V_{gs} = \frac{1/j\omega C_{\text{eq}}}{(R_s \| R_G) + 1/j\omega C_{\text{eq}}} \mathbf{V}_s = \frac{\mathbf{V}_s}{1 + j\omega C_{\text{eq}}(R_s \| R_G)} \quad (18\text{-}25)$$

As frequency is increased, a smaller fraction of the voltage V_s appears across C_{eq} and voltage V_{gs} is reduced. At the upper cutoff frequency defined by

$$\omega_{21} = \frac{1}{C_{\text{eq}}(R_s \| R_G)} \quad (18\text{-}26)$$

the input voltage V_{gs} will be down to 70% of V_s and the possible amplifier gain will be correspondingly reduced. Reasoning by analogy, we can say that at an upper cutoff frequency

$$\omega_{22} = \frac{1}{C_o(R_D \| R_L)} \quad (18\text{-}27)$$

the output voltage V_o will be down to 70% of $g_m V_{gs} R_o$ and the amplifier gain will be correspondingly reduced. The overall high-frequency gain for the FET amplifier can be expressed as

$$\mathbf{A}_H = \mathbf{A}_O \cdot \frac{1}{1 + j\omega C_{\text{eq}}(R_s \| R_G)} \cdot \frac{1}{1 + j\omega C_o(R_D \| R_L)} \tag{18-28}$$

or the relative gain at high frequencies is

$$\frac{\mathbf{A}_H}{\mathbf{A}_O} = \frac{1}{1 + j\omega/\omega_{21}} \cdot \frac{1}{1 + j\omega/\omega_{22}} \tag{18-29}$$

To predict the behavior of a given circuit, determine ω_{21} and ω_{22} from Eqs. 18-26 and 27; the lower of the two determines the upper cutoff frequency (unless they are very close together). In designing a circuit for a specified behavior, the designer does not have freedom in specifying C_{eq} and C_o since these are determined by device parameters, operating conditions, and gain. If the predicted performance is not satisfactory, one possibility is to reduce R_D which reduces gain (and therefore C_{eq}) and increases ω_{21} and ω_{22}.

Example 4

In Example 1, an FET ($g_m = 2000\ \mu\mho$, $C_{gs} = 20$ pF, and $C_{gd} = 2$ pF) was used in an RC-coupled amplifier with $R_s = 1\ \text{k}\Omega$, $R_G = 1\ \text{M}\Omega$, $R_D = 5\ \text{k}\Omega$, and $R_L = 5\ \text{k}\Omega$ to give a midfrequency gain of $-g_m R_o = -5$. If $C_W + C_L \cong 4 + 16 = 20$ pF, predict the upper cutoff frequency.

The equivalent input and output capacitances are

$$C_{\text{eq}} = C_{gs} + C_{gd}(1 + g_m R_o) = 20 + 2(6) = 32\ \text{pF}$$
$$C_o = C_W + C_L + C_{gd} = 4 + 16 + 2 = 22\ \text{pF}$$

The equivalent input and output resistances are

$$R_s \| R_G = 10^3 \times 10^6/(10^3 + 10^6) = 999 \cong 1000\ \Omega$$
$$R_D \| R_L = 5000 \times 5000/(5 + 5) \times 10^3 = 2500\ \Omega$$

The individual cutoff frequencies are

$$\omega_{21} = 1/C_{\text{eq}}(R_s \| R_G) = 1/32 \times 10^{-12} \times 10^3 = 31.3\ \text{Mrad/s}$$

$$\omega_{22} = 1/C_o(R_D \| R_L) = 1/22 \times 10^{-12} \times 2500 = 18\ \text{Mrad/s}$$

The upper cutoff frequency of the amplifier is

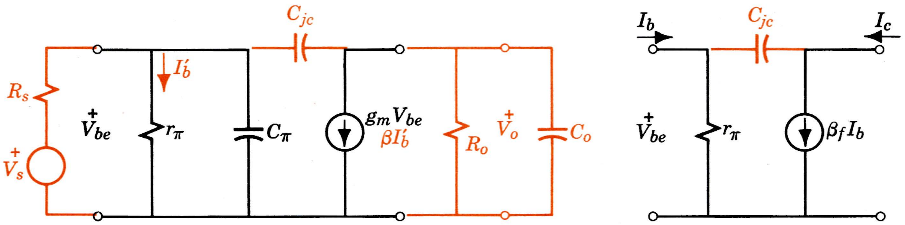
$$18/2\pi \cong 2.9\ \text{MHz}$$

The manufacturer's data sheets usually give values that are easy to measure. The "reverse transfer capacitance" $C_{rss} = C_{gd}$; the "input capacitance with output shorted" $C_{iss} = C_{gs} + C_{gd}$. Therefore, $C_{gs} = C_{iss} - C_{rss}$. We have not considered C_{ds}, which is usually less than 1 pF.

BJT. Comparing Figs. 18.12a and b and reasoning by analogy, the upper cutoff frequency for a BJT amplifier is defined by

$$\omega_{21} = \frac{1}{C_{\text{eq}}(R_s \| r_\pi)} \quad \text{or} \quad \omega_{22} = \frac{1}{C_o(R_C \| R_L)} \tag{18-30}$$

where $C_{\text{eq}} = C_\pi + C_{jc}(1 + g_m R_o)$ and $C_o = C_{jc} + C_W + C_L$ (Fig. 18.15a). (More precisely, the equivalent input resistance is $R_s \| r_\pi \| R_B$, but in a well-designed amplifier $R_B \gg r_\pi$.) The overall high-frequency gain can be expressed as



(a) Hybrid-π model (b) Frequency-dependent β

Figure 18.15 High-frequency BJT models.

$$\mathbf{A}_H = \mathbf{A}_O \cdot \frac{1}{1 + j\omega C_{\text{eq}}(R_s \| r_\pi)} \cdot \frac{1}{1 + j\omega C_o(R_C \| R_L)} \tag{18-31}$$

and the relative gain at high frequencies is

$$\frac{\mathbf{A}_H}{\mathbf{A}_O} = \frac{1}{1 + j\omega/\omega_{21}} \cdot \frac{1}{1 + j\omega/\omega_{22}} \tag{18-32}$$

The upper cutoff frequency is determined by the lower of ω_{21} and ω_{22} just as in the case of the FET amplifier.

Analysis of the high-frequency response of transistors is complicated by the fact that *lumped* circuit models can only approximate effects that are, in fact, *distributed* throughout a finite region in the transistor. For example, finite times are required for signals in the form of changes in charge density to be propagated across the base of a bipolar transistor. At higher frequencies the current gain I_c/I_b begins to decrease in magnitude and a phase lag appears. In the hybrid-π model, these charging effects are represented by capacitance C_π.

Another approach is to represent the transistor by parameters that are frequency dependent. We define β as the short-circuit current gain at low and medium frequencies and let $I_c/I_b = \beta_f$, a function of frequency. At the *beta cutoff frequency* f_β, current gain β_f is just equal to 70.7% of the low-frequency value. The significant relation between current phasors (Fig. 18.15b) is

$$\frac{\mathbf{I}_c}{\mathbf{I}_b} = \frac{\beta}{1 + j(f/f_\beta)} = \frac{\beta}{1 + j(\omega/\omega_\beta)} \tag{18-33}$$

and the magnitude of the high-frequency current gain is

$$\beta_f = \frac{\beta}{\sqrt{1 + (f/f_\beta)^2}} \tag{18-34}$$

where f_β may be specified by the manufacturer. At very high frequencies, $\beta_f = \beta/(f/f_\beta) = \beta f_\beta/f$ or current gain is inversely proportional to frequency. Alternatively, therefore, the manufacturer may specify the frequency

$$f_T = \beta f_\beta \tag{18-35}$$

at which the current gain β_f would fall to unity, or he may state the value of β_f at a specified frequency f.

To avoid working with frequency-dependent parameters, we can use constant circuit elements with the appropriate frequency response. In Fig. 18.15a, the fall-off in short-circuit current gain is accounted for by C_π alone since A is zero. Therefore,

$$f_T = \beta f_\beta = \frac{\beta}{2\pi C_\pi r_\pi} = \frac{1}{2\pi C_\pi r_e} \tag{18-36}$$

since $r_\pi = \beta r_e$. The determination of parameter values from manufacturer's specifications is illustrated in Example 5.

Example 5

The manufacturer of the 2N3114 specifies $h_{ib} = r_e = 27$ and $h_{fe} = \beta = 50$ at 1 kHz (Appendix 3) and $\beta = 2.7$ at 20 MHz. Determine f_T, f_β, and C_π.

The 2.7 value is for β_f at $f = 20$ MHz. By Eq. 18-35, the current gain becomes unity at $f_T = \beta f_\beta = \beta_f \times f = 2.7 \times 20 \text{ MHz} = 54 \text{ MHz}$ and the beta cut-off frequency is

$$f_\beta = \frac{f_T}{\beta} = \frac{54 \text{ MHz}}{50} = 1.08 \text{ MHz}$$

Since the common-base emitter resistance $h_{ib} = r_e$, by Eq. 18-36,

$$C_\pi = \frac{1}{2\pi f_T r_e} = \frac{1}{2\pi \times 54 \times 10^6 \times 27} = 109 \text{ pF}$$

For ordinary transistors used in audiofrequency amplifiers, the decrease in β is unimportant; for special high-frequency planar transistors, the effect may be unimportant up to many megahertz. In amplifiers employing high-frequency transistors, the upper cutoff frequency may be determined by the effective shunting capacitance C_o, due to wiring and other parasitic effects, as indicated by Eq. 18-30. If the voltage gain is large, the collector-base junction capacitance C_{jc} becomes critical. (The practical value of this parameter is the common-base output capacitance designated C_{ob} or C_{cb} on the data sheet.) In practice, the engineer determines which is the limiting factor and bases his amplifier design on a careful analysis of that factor.

Multistage Amplifiers

To achieve the desired voltage or current gain and the necessary frequency response, several stages of amplification may be required.

Cascading

The stages are usually connected in *cascade;* the output of one stage is connected to the input of the next. Where the preceding and following stages are similar, $R_L = r_\pi$ (Fig. 18.5b) and the midfrequency *per stage* voltage gain for the BJT is

$$A_{VO} = \frac{V_o}{V_{be}} = \frac{-g_m V_{be} R_o}{V_{be}} = -g_m R_o \tag{18-37}$$

just as for the FET. For the BJT the midfrequency *per stage* current gain is

$$A_{IO} = \frac{I_o}{I_b} = \frac{V_o/r_\pi}{V_{be}/r_\pi} = \frac{V_o}{V_{be}} = -g_m R_o = -\frac{\beta}{r_\pi} R_o \tag{18-38}$$

Since R_o is determined in part by r_π and is always less than r_π, the current gain per stage is always less than β.

By defining a single stage to include the input characteristics of the next amplifying element, the combined effect of cascaded stages is easy to predict. In Fig. 18.16, the individual stages are represented by rectangular blocks labeled with the individual voltage gains. The overall voltage gain is

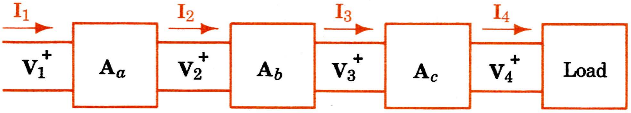


Figure 18.16 Block diagram of amplifiers in cascade.

$$\mathbf{A} = \frac{\mathbf{V}_4}{\mathbf{V}_1} = \frac{\mathbf{V}_2}{\mathbf{V}_1}\frac{\mathbf{V}_3}{\mathbf{V}_2}\frac{\mathbf{V}_4}{\mathbf{V}_3} = \mathbf{A}_a\mathbf{A}_b\mathbf{A}_c \tag{18-39}$$

or

$$A\underline{/\theta} = A_a A_b A_c \underline{/\theta_a + \theta_b + \theta_c} \tag{18-40}$$

In words, the overall voltage gain is the product of the gain amplitudes and the sum of the phase shifts; both A and θ are functions of frequency. A similar statement would describe the overall current gain.

Gain in Decibels

The calculation of overall gain in multistage amplifiers is simplified by using decibels. As indicated in Eq. 9-27,

$$\text{Gain in decibels} = 20 \log \frac{V_2}{V_1} \tag{18-41}$$

The advantage of the decibel unit here is that when response is plotted in dB the overall response curves of a multistage amplifier can be obtained by adding the individual response curves. In Fig. 18.17, the response curves for two stages a and b are plotted in dB. From Eqs. 18-13 and 18-32 it can be deduced that for $f \ll f_1$ relative gain is proportional to frequency, and for $f \gg f_2$ relative gain is inversely proportional to frequency. On log-log scales, these relations are straight lines with slopes of 20 dB per decade and the response of an RC-coupled amplifier can be approximated by curve A_a. (Actually, the gain is down 3 dB at the *break-point frequencies.*) A second amplifier with better low-frequency response is approximated by curve A_b. The overall gain of the combination is the sum of the two curves, and the straight-line approximation is easily drawn. For a single-stage amplifier with two high-frequency breakpoints, the frequency response would resemble curve A of Fig. 18.17.

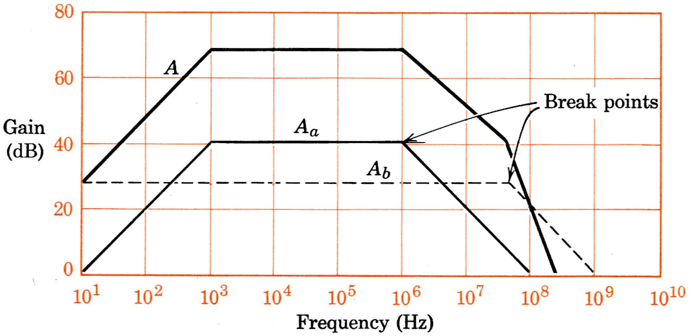


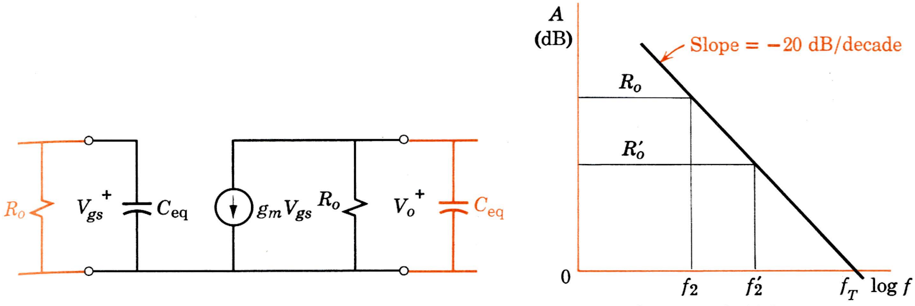
Figure 18.17 Frequency response expressed in decibels.

Gain-Bandwidth Product

The *bandwidth* of an amplifier is a useful design criterion; an audio amplifier needs a bandwidth of about 20 kHz whereas a video amplifier requires a bandwidth of about 4 MHz. For an untuned amplifier the bandwidth is defined as $\text{BW} = \omega_2 - \omega_1 \cong \omega_2$ since ω_1 is small compared to ω_2. For a single-stage (Fig. 18.18), the gain is

$$A = \frac{V_o}{V_{gs}} = \frac{g_m}{\dfrac{1}{R_o} + j\omega C_{\text{eq}}} = \frac{g_m R_o}{1 + j\omega R_o C_{\text{eq}}} = \frac{A_o}{1 + j\omega R_o C_{\text{eq}}} \tag{18-42}$$

Where $\omega_2 = 1/R_o C_{\text{eq}}$, any increase in bandwidth by decreasing R_o results in a



(a) FET stage (b) Gain-bandwidth relation

Figure 18.18 Gain-bandwidth relation for an FET amplifier.

decrease in midfrequency gain. The product of gain magnitude and bandwidth is

$$A_o\omega_2 = g_m R_o \frac{1}{R_o C_{eq}} = \frac{g_m}{C_{eq}} \leq \frac{g_m}{C_{gs}} \tag{18-43}$$

since $C_{eq} \cong C_{gs}$ if the Miller effect is not large. Because g_m and C_{gs} are device parameters, the *gain-bandwidth product* is a constant for a specific FET. For the FET of Example 4, $g_m/C_{gs} = 2 \times 10^{-3}/20 \times 10^{-12} = 10^8$ rad/s $\cong$ 15 MHz. With this FET a gain of 5 permits a bandwidth of less than 3 MHz (2.9 MHz in Example 4 where $R_s = 1\ \text{k}\Omega$); if a bandwidth of 5 MHz is required, the maximum possible gain is 3.

A bipolar transistor has a similar *figure of merit,* but it is usually expressed in a different way. If the high-frequency response is limited by the beta cutoff, the bandwidth is approximately f_β and the maximum possible gain is β, so the *gain-bandwidth product* is

$$A_o f_2 \leq \beta f_\beta = f_T \tag{18-44}$$

The frequency f_T at which $\beta_f = 1$ typically ranges from 100 kHz for alloy junction power transistors to 10 GHz for high-frequency transistors.

Example 6

The specifications for a certain silicon transistor indicate a minimum h_{fe} of 40 and a typical f_T of 8 MHz. Is this transistor suitable as a radiofrequency amplifier at 1 MHz?

By Eq. 18-35 at $f = 1$ MHz,

$$\beta_f = \frac{\beta f_\beta}{f} = \frac{f_T}{f} = \frac{8}{1} = 8$$

By Eq. 18-44,

$$f_\beta = \frac{f_T}{\beta} = \frac{f_T}{h_{fe}} = \frac{8\ \text{MHz}}{40} = 0.2\ \text{MHz}$$

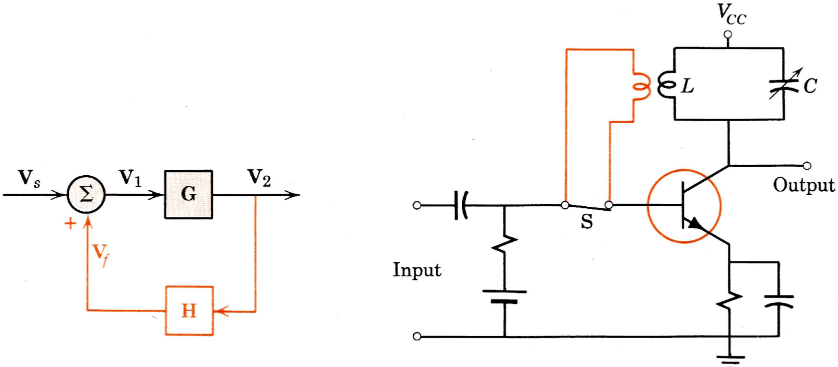
Although some current gain is possible at 1 MHz, the beta cutoff is only 0.2 MHz and another transistor should be selected.

Feedback Amplifiers

The concept of feedback was introduced in Chapter 9 and was used in Chapter 16 to stabilize the operating points of field-effect and bipolar transistors. The combination of amplifier gain and positive or negative feedback can be used to perform some very valuable functions.

Positive Feedback

By using the rules of block diagram algebra, all linear feedback networks can be reduced to the basic system of Fig. 18.19a, for which the gain is $\mathbf{G}_F = \mathbf{G}/(1 - \mathbf{GH})$. The behavior of the system can be predicted by examining the denominator $1 - \mathbf{GH}$.



(a) Basic feedback system (b) Resonant-circuit oscillator

Figure 18.19 Positive feedback in an oscillator circuit.

If the value of **GH** approaches $1 + j0$, the denominator approaches zero and the overall gain increases without limit. An amplifier of infinite gain can provide an output with no input; such a device is called an *oscillator* or signal generator and is an important part of many electronic systems.

Oscillators

When a circuit is designed to oscillate,† special provision is made to feed back a portion of the output signal in the correct phase and amplitude to make the **GH** product just equal to $1 + j0$. By making **GH** frequency dependent, oscillation can be obtained at a single desired frequency. Either **G** or **H** can be the frequency-dependent element of the feedback system.

---

† The fact that stable operation occurs in the nonlinear region of the device characteristic complicates the design and analysis of oscillators.

The oscillator of Fig. 18.19b is basically an amplifier with a parallel resonant circuit for a load impedance. Gain is a maximum for signals at the resonant frequency of the collector circuit. This is called a *tuned* amplifier because the frequency for maximum amplification can be selected by varying the capacitance. (Such an amplifier is used in a radio receiver to amplify the signals transmitted by a selected station while rejecting all others.) To create a "self-excited amplifier" or oscillator, switch S is opened (and the "Input" terminals are shorted) and part of the voltage across the tuned circuit is fed back into the input. In this case, **G** is a sensitive function of frequency.

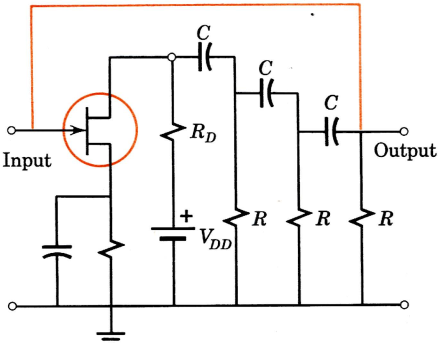
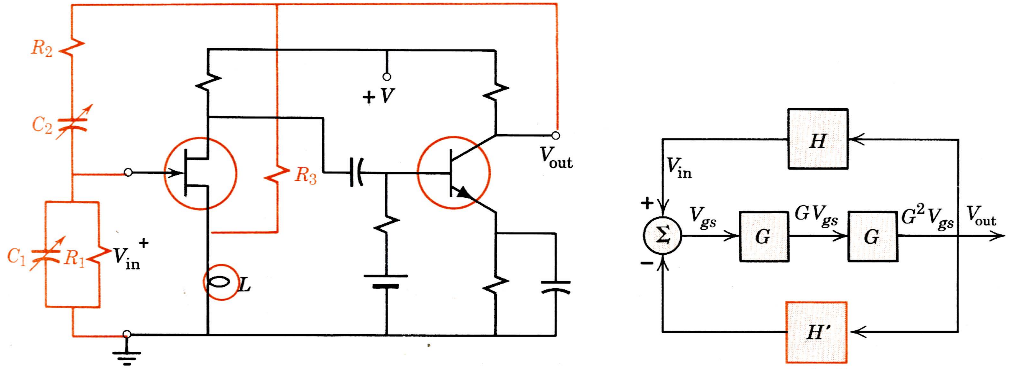


Figure 18.20 An FET phase-shift oscillator.

For audiofrequencies, the values of L and C required for a resonant circuit are impractically large. The *phase-shift oscillator* of Fig. 18.20 avoids this requirement and is fairly common for the generation of audiofrequency signals. The output of the FET amplifier appears across R_D. The first RC section constitutes a phase-shifting network in that the voltage across R is from 0° to 90° ahead of the voltage across R_D depending on the frequency. For a particular frequency, the three RC sections produce a total phase shift of 180°; the additional 180° required for oscillation is inherent in the gate-drain voltage relations in an FET. In this case, **H** is the frequency-sensitive element of the feedback system.

To provide good waveform, an oscillator should operate in the linear region of the amplifier. In the *resistance-capacitance tuned oscillator* of Fig. 18.21a,



(a) Circuit diagram

(b) Block diagram

Figure 18.21 A resistance-capacitance tuned oscillator.

widely used as a laboratory instrument, the necessary phase reversal for positive feedback is provided by the two-stage amplifier. The input to the first stage is obtained from the voltage-divider $Z_1/(Z_1 + Z_2)$ across the output of the second stage. Voltage V_{in} is in phase with V_{out} for $\omega = 1/\sqrt{R_1R_2C_1C_2}$ (the principle of the Wien bridge) and oscillation occurs. In the usual case, $R_1 = R_2 = R$ and $C_1 = C_2 = C$ and the oscillator frequency is $\omega_o = 1/RC$. As the oscillations build up, the current through R_3 and incandescent lamp L increases, raising the temperature of the filament. The voltage developed across this nonlinear resistance subtracts from the input voltage and limits the amplitude of the oscillations. The negative feedback loop is shown in Fig. 18.21b, where H' is a nonlinear function of output voltage.

Negative Feedback

Negative feedback, where the amplitude and phase of the signal fed back are such that the overall gain of the amplifier is less than the forward gain, may be used to advantage in amplifier design. Since additional voltage or current gain can be obtained easily with transistors or tubes, the loss in gain is not important if significant advantages are obtained. In addition to stabilizing gains and reducing time constants (see p. 238), negative feedback can be used to improve amplifier performance in the following ways:

1. Operating point can be made practically independent of device parameters, and can be made insensitive to temperature changes or aging of components.
2. Gain can be made practically independent of reactive elements and therefore insensitive to frequency.
3. Gain can be made selective to discriminate against distortion or noise.
4. The input and output impedance of an amplifier can be improved greatly.

Stabilizing Operation

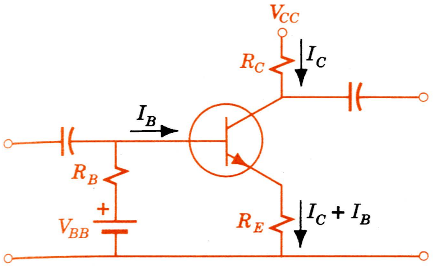
The self-biasing circuit for decreasing the effect of changes in β or temperature on the operating point of a transistor is essentially a feedback device. As indicated in Fig. 18.22a, if I_C tends to increase, the current $I_C + I_B$ in R_E increases, raising the potential of the emitter with respect to ground. This, in turn, reduces the forward bias, reduces the base current and, therefore, limits the increase in I_C. In other words, an increase in I_C is fed back in such a way as to oppose the change that produced it. (See Example 7.)

Improving Frequency Response

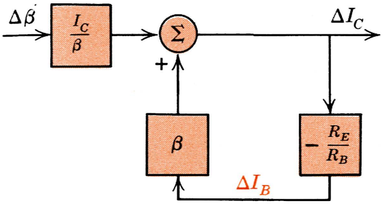
No amplifier amplifies equally well at all frequencies; at the upper cutoff frequency, for example, the voltage gain is down to 70.7% of the midfrequency value. By sacrificing gain, however, the upper cutoff frequency can be raised.

Example 7

Represent the response of the self-biasing circuit of Fig. 18.22a by a block diagram and derive the "stability factor" $\Delta I_C/\Delta\beta$.



(a) Self-biasing circuit



(b) Block diagram

Figure 18.22 Bias stabilization.

Where I_{CBO} is negligible (as in silicon transistors), $I_C = \beta I_B$ and

$$\Delta I_C = \Delta\beta I_B + \beta\,\Delta I_B = \Delta\beta(I_C/\beta) + \beta\,\Delta I_B \tag{18-45}$$

With fixed-current bias ($\Delta I_B = 0$), changes in β produce proportionate changes in I_C.

With self-bias (Fig. 18.22a),

$$V_{BB} - I_B R_B - V_{BE} - (I_C + I_B)R_E = 0$$

Assuming V_{BE} is constant, for small changes, this becomes

$$-\Delta I_B R_B - \Delta I_C R_E - \Delta I_B R_E = 0$$

or

$$\Delta I_B = -\Delta I_C R_E/(R_B + R_E) \cong -\Delta I_C R_E/R_B \tag{18-46}$$

where R_B is large compared to R_E.

Equations 18-45 and 46 are realized in Fig. 18.22b. For this closed-loop system $G = 1$, and

$$\frac{\Delta I_C}{\Delta\beta} = \frac{I_C/\beta}{1 + \beta R_E/R_B} \quad \text{or} \quad \frac{\Delta I_C}{I_C} = \frac{\Delta\beta/\beta}{1 + \beta R_E/R_B}$$

For $\beta R_E = 10R_B$ (Eq. 16-18), an 11% change in β results in only a 1% change in I_C.

In the practical case, gain can be made insensitive to frequency over the range of interest by using negative feedback.

Using system notation, the gain **G** is frequency dependent because of storage effects in the electronic device and reactive elements in the circuit. The feedback factor **H,** however, can be obtained through purely resistive elements. If **H** is obtained by means of a voltage divider tap on a precision noninductive resistor, then $\mathbf{H} = H$, a constant. If $\mathbf{G}H$ is very large at the frequency of interest, the basic gain equation becomes

$$\mathbf{G}_F = \frac{\mathbf{G}}{1 - \mathbf{G}H} \cong \frac{\mathbf{G}}{-\mathbf{G}H} = -\frac{1}{H} \tag{18-47}$$

and the gain is independent of frequency, temperature, or aging.

Example 8

An amplifier whose high-frequency response is given by $\mathbf{G} = A_o/[1 + j(\omega/\omega_2)]$ is used in the basic feedback circuit (Fig. 18.19a). If $A_o = 1000$, $\omega_2 = 10^4$ rad/s, and $H = -0.009$, derive an expression for high-frequency response with feedback and predict the new upper cutoff frequency ω_{2F}.

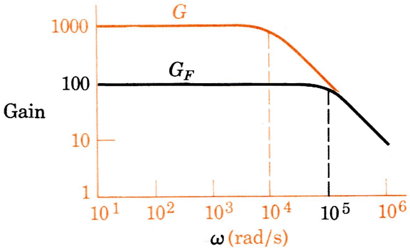


Figure 18.23 Gain-bandwidth tradeoff.

The response is shown in Fig. 18.23. With feedback,

$$\mathbf{G}_F = \frac{\mathbf{G}}{1 - \mathbf{GH}} = \frac{\dfrac{1000}{1 + j(\omega/\omega_2)}}{1 + \dfrac{9}{1 + j(\omega/\omega_2)}}$$

$$= \frac{1000}{1 + j\left(\dfrac{\omega}{\omega_2}\right) + 9} = \frac{100}{1 + j\left(\dfrac{\omega}{10\omega_2}\right)} \qquad (18\text{-}48)$$

In the midfrequency range, ω is small compared to $10\omega_2$ and

$$G_F \cong 100$$

At the new cutoff frequency, gain will be down by a factor of $\sqrt{2}$; therefore, $\omega_{2F}/10\omega_2 = 1$, or

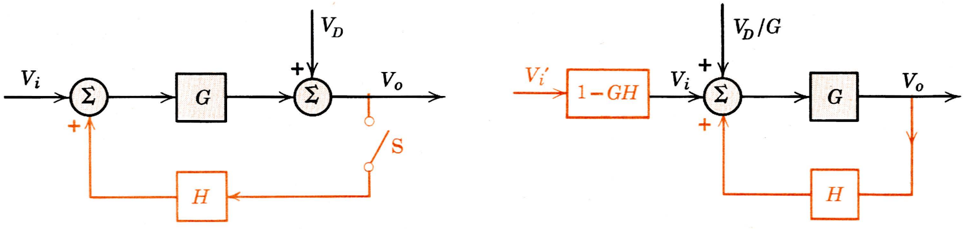
$$\omega_{2F} = 10\omega_2 = 10(10^4) = 10^5 \text{ rad/s}$$

In Example 8, the midfrequency gain is decreased by a factor of $1 - GH$, but the bandwidth has been increased by the same factor. As is usual in such cases, we have traded gain for bandwidth.

Reducing Distortion

Distortion-free amplification is obtained easily at low signal levels. In the last power amplifier stage, however, efficient performance usually introduces nonlinear distortion because of the large signal swings. A common type of distortion results in the introduction of a second harmonic as shown in Fig. 16.1. (Also see Eq. 17-3.)

A nonlinear amplifier can be represented by the combination of linear blocks and rms signals shown in Fig. 18.24a with switch S open. The distortion compo-



(a)

(b)

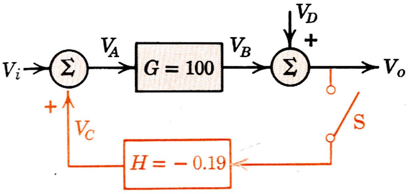
Figure 18.24 Block diagram representation of a distortion component.

nent V_D is introduced in the output with no corresponding component in the input V_i. Without feedback, the output signal is

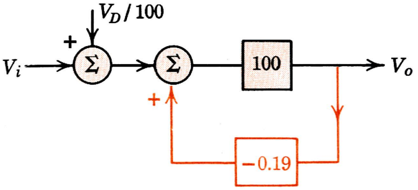
$$V_o = GV_i + V_D \tag{18-49}$$

Example 9

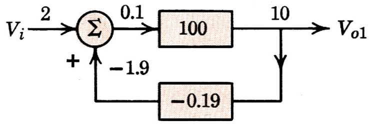
The output of a power amplifier contains a 10-V, 1000-Hz fundamental and "20% second-harmonic distortion." The gain $\mathbf{G} = 100\angle 0°$. Predict the effect of feedback on the distortion if $H = -0.19$.



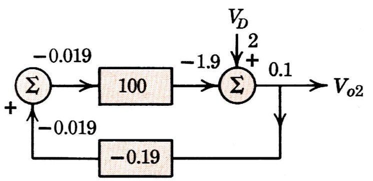
(a) Power amplifier



(b) Modified design



(c) 1000-Hz signal



(d) 2000-Hz signal

Figure 18.25 Reducing distortion in a power amplifier by negative feedback.

The block diagram for the system is as shown in Fig. 18.25a. Blocks G and H are linear models so the distortion component V_D is introduced as an extraneous signal not present in the input to the amplifier. By definition, 20% second-harmonic distortion indicates the presence of a 2000-Hz signal of magnitude

$$V_D = 0.20V_{B1} = 0.20 \times 10 = 2\text{ V}$$

Applying rule 4 (p. 235), the summing point for V_D is shifted in front of G and the block diagram is that of Fig. 18.25b. With feedback, the gain is

$$G_F = \frac{G}{1 - GH} = \frac{100}{1 + 19} = 5$$

To retain the original power output, the input signal must be increased to

$$V_i = \frac{V_{o1}}{G_F} = \frac{10}{5} = 2\text{ V at 1000 Hz}$$

The output is now G_F times the equivalent input or

$$V_o = G_F(V_i + V_D/100)$$

$$\text{At 1000 Hz,}\quad V_{o1} = G_F V_i = 5 \times 2 = 10\text{ V}$$

$$\text{At 2000 Hz,}\quad V_{o2} = G_F V_D/100 = 5 \times 0.02 = 0.1\text{ V}$$

The voltage distributions for fundamental and second-harmonic signals are shown in Figs. 18.25c and d. Since the blocks are linear, the superposition principle is applicable and the two signals can be handled separately.

The percentage of second-harmonic distortion is now

$$\frac{V_{o2}}{V_{o1}} \times 100 = \frac{0.1}{10} \times 100 = 1\%\ \text{(a tolerable amount)}$$

The original amplifier required an input signal of $V_{o1}/100 = 0.1$ V; therefore, additional, distortionless, voltage gain of $2\text{ V}/0.1\text{ V} = 20$ is necessary, but this is easily obtainable at these low levels.

and the distortion in percent of the fundamental is $(V_D/GV_i) \times 100$. Closing the switch S introduces feedback. In Fig. 18.24b, using the corollary of rule 4 on p. 235, the distortion component V_D has been shifted ahead of G by inserting the factor $1/G$. Now

$$V_o = \frac{G(V_i + V_D/G)}{1 - GH} = \frac{G}{1 - GH} V_i + \frac{V_D}{1 - GH}$$

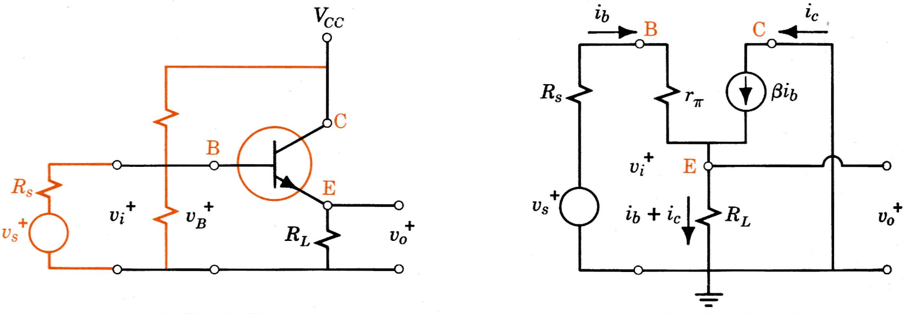
If another stage of low-level, distortionless amplification with a gain of $1 - GH$ is added, the new output is

$$V_o = GV_i' + \frac{V_D}{1 - GH} \tag{18-50}$$

The relative amplitude of the distortion component compared to the desired component has been reduced by the factor $1 - GH$, which may be 10 to 100 or so. (See Example 9 on preceding page.)

Improving Impedance Characteristics

To minimize the "loading" effect of an electronic device such as an amplifier on a signal source, the input impedance of the device should be high. An FET with insulated gate is very good in this respect, but a BJT is not. Also, to provide efficient power transfer to the load, the output impedance of the device should be small. Neither the FET nor the BJT is optimum in this respect. By using negative feedback the input and output impedances can be dramatically improved. A particularly useful form of feedback circuit is the FET *source follower* or its BJT counterpart, the *emitter follower*.



(a) Circuit diagram (b) Circuit model

Figure 18.26 The emitter follower, a common-collector amplifier with feedback.

Followers

In the circuit of Fig. 18.26a, the load resistor R_L is in series with the emitter and the collector is tied to V_{CC}, an ac ground. The output voltage† $v_o = v_E$ differs from the base voltage v_B by the base-emitter voltage $V_{BE} \cong 0.7$ V. Even for large swings of v_B, the voltage of the emitter with respect to ground "follows"; hence the name *emitter follower.*

---

† In this discussion we use lower-case letters to represent small signals. The relations hold equally well for effective values of sinusoidal signals.

The Emitter Follower

To examine the impedance characteristics of the emitter follower at moderate frequencies, we replace the transistor by its small-signal model (Fig. 18.26b). Summing voltages around the input loop and solving for the input current,

$$i_b = \frac{v_s}{R_s + r_\pi + (\beta + 1)R_L} \tag{18-51}$$

The output voltage is

$$v_o = (i_b + i_c)R_L = (\beta + 1)i_b R_L = \frac{(\beta + 1)R_L v_s}{R_s + r_\pi + (\beta + 1)R_L} \tag{18-52}$$

and the voltage gain in this feedback circuit is

$$G_F = \frac{v_o}{v_s} = \frac{(\beta + 1)R_L}{R_s + r_\pi + (\beta + 1)R_L} \tag{18-53}$$

If β is large (50 to 150) and if R_L is of the same order of magnitude as $(R_s + r_\pi)$, then G_F approaches unity, but it can never exceed unity.

The virtue of an emitter follower is not in providing voltage gain. It does provide a current gain of $\beta + 1$ and a corresponding power gain, but its greatest usefulness is in providing impedance transformation. From Fig. 18.26b,

$$v_i = i_b r_\pi + (i_b + i_c)R_L = i_b r_\pi + (\beta + 1)i_b R_L \tag{18-54}$$

By definition, the input resistance is

$$R_i = \frac{v_i}{i_b} = r_\pi + (\beta + 1)R_L \cong (\beta + 1)R_L \tag{18-55}$$

The effective output resistance can be evaluated from open- and short-circuit conditions. With the output terminals open-circuited,

$$v_{OC} = v_o \cong v_s$$

since the gain is approximately unity (Eq. 18-53). With the output terminals shorted (shorting R_L), $i_b = v_s/(R_s + r_\pi)$ and the short-circuit current is

$$i_{SC} = i_b + i_c = (\beta + 1)i_b = \frac{\beta + 1}{R_s + r_\pi} v_s$$

Therefore,

$$R_o = \frac{v_{OC}}{i_{SC}} = \frac{R_s + r_\pi}{\beta + 1} \tag{18-56}$$

From Eqs. 18-55 and 18-56, it is clear that the emitter follower acts as an impedance transformer with a ratio of transformation approximately equal to $\beta + 1$. With a load resistance R_L, the resistance seen at the input terminals is $(\beta + 1)$ times as great. With the input circuit resistance $R_s + r_\pi$, the effective output resistance is only $1/(\beta + 1)$ times as large.

Example 10

A 2N1893 transistor ($r_\pi = 2.8\ \text{k}\Omega$, $\beta = 70$) is used as an amplifier with $R_s = 1.2\ \text{k}\Omega$ and $R_L = 4\ \text{k}\Omega$. Compare the input and output resistances in the common-emitter and common-collector configurations.

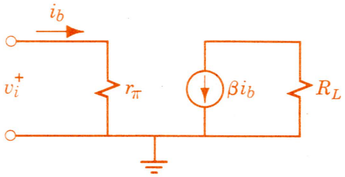


Figure 18.27 Calculation of input resistance

Using the simplified circuit of Fig. 18.27, the input resistance in the common-emitter configuration is

$$R_i = \frac{v_i}{i_b} = r_\pi = 2.8\ \text{k}\Omega$$

For the common-collector circuit of Fig. 18.26a, the input resistance (Eq. 18-55) is

$$R_i = r_\pi + (\beta + 1)R_L = 2.8 + (70 + 1)4 = 287\ \text{k}\Omega$$

The *CC* input resistance is over 100 times as great.

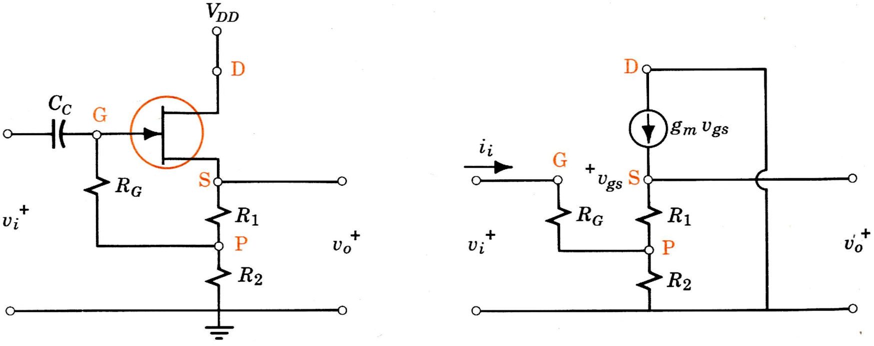
In the *CE* amplifier with $R_C = 4\ \text{k}\Omega$, $R_o = 4\ \text{k}\Omega$. In the *CC* amplifier,

$$R_o = \frac{R_s + r_\pi}{\beta + 1} = \frac{1200 + 2800}{70 + 1} = 56\ \Omega$$

The output resistance is reduced by a factor of 71.

The Source Follower

The common-drain amplifier or "source follower" is similar to the emitter follower in that it provides improved (lower) output impedance in return for reduced (essentially unity) gain. The input impedance, already high for an FET, can be made even higher by using the bias arrangement of Fig. 18.28 where



(a) Circuit diagram (b) Small-signal model

Figure 18.28 The source follower or common-drain amplifier.

$R_S = R_1 + R_2$. The quiescent operating point is defined by

$$V_{DS} = V_{DD} - I_D(R_1 + R_2) \tag{18-57}$$

where V_{DS} is approximately half of V_{DD} to put operation near the midpoint of the load line. The gate-source bias,

$$V_{GS} = -I_D R_1 \tag{18-58}$$

is typically a few volts so that $R_1 \ll R_2$ and $R_S \cong R_2$.

To evaluate the performance of the source follower, we replace the FET by its small-signal model (Fig. 18.28b). In this circuit, $v_{gs} = v_i - v_o$; in feedback notation, $H = 1$ and $G = g_m R_S$. Neglecting the current in the very large R_G and equating the currents into node S, $v_o/R_S = g_m v_{gs} = g_m(v_i - v_o)$ or

$$v_o = \frac{g_m R_S v_i}{1 + g_m R_S} \tag{18-59}$$

and the voltage gain is

$$G_F = \frac{v_o}{v_i} = \frac{g_m R_S}{1 + g_m R_S} \tag{18-60}$$

which approaches unity in the typical case where $g_m R_S \gg 1$.

The effective output resistance can be evaluated from open- and short-circuited conditions. Since the open-circuit gain is approximately unity,

$$v_{OC} = v_o \cong v_i$$

With the output terminals shorted (shorting R_S),

$$i_{SC} = g_m v_{gs} = g_m(v_i - 0) = g_m v_i$$

Therefore, the effective output resistance is

$$R_o = \frac{v_{OC}}{i_{SC}} \cong \frac{v_i}{g_m v_i} \cong \frac{1}{g_m} \tag{18-61}$$

For a typical g_m of 2000 μ℧, the output resistance is about 500 Ω as compared to a common-source value (R_D) of several thousand ohms.

The higher input resistance resulting from the bias circuit of Fig. 18.28a can be anticipated in a qualitative way as follows. The ac input resistance is the ratio of an increment of voltage to the corresponding increment of current. An increment of voltage at point G produces an equal increment of voltage at point S and, since $R_1 \ll R_2$, a nearly equal increment of voltage at point P. Therefore, the voltage across R_G, and the corresponding current through R_G, changes very little and the input resistance is much higher than if the lower end of R_G were returned to ground.

For a quantitative analysis, we return to Fig. 18.28b. Neglecting the effect of the very small current i_i in R_2, the voltage at point P is

$$v_P = \frac{R_2}{R_1 + R_2} v_o = \frac{R_2}{R_1 + R_2} G_F v_i \tag{18-62}$$

Then the effective input resistance is

$$R_i = \frac{v_i}{i_i} = \frac{v_i}{(v_i - v_P)/R_G} = \frac{R_G}{1 - \dfrac{R_2}{R_1 + R_2} G_F} \tag{18-63}$$

In the typical case where $R_2 \cong R_1 + R_2$ and $G_F \cong 1$, the input resistance is many megohms and the source follower presents a negligible load to the signal source. This characteristic makes it desirable as the input stage of a multistage amplifier. (*Note:* At signal frequencies, the input capacitance $C_i \cong C_{gd}$ in parallel with R_i may determine the input impedance. See Problem 9.)

SUMMARY

◆ Linear models are used to predict the performance of amplifiers in which signals are much smaller than bias values.

The frequency-response curves of all *RC*-coupled amplifiers are similar in shape.

◆ For one stage of *RC* coupled voltage (or current) amplification,

Midfrequency gain is $A_O = -g_m R_o$ (FET) or $-\beta R_o/(r_\pi + R_s)$ (BJT)

Low-frequency response is $\mathbf{A}_L = A_O \dfrac{1}{1 - j(\omega_{11}/\omega)} \cdot \dfrac{1}{1 - j(\omega_{12}/\omega)}$

High-frequency response is $\mathbf{A}_H = A_O \dfrac{1}{1 + j(\omega/\omega_{21})} \cdot \dfrac{1}{1 + j(\omega/\omega_{22})}$

Lower and upper cutoff frequencies ω_1 and ω_2 are determined by the critical combinations of *R* and *C*.

◆ For two stages of amplification in cascade, the overall gain is

$$A\underline{/\theta} = A_1 A_2 \underline{/\theta_1 + \theta_2}$$

Gain in dB = $10 \log (P_2/P_1)$ or $20 \log (V_2/V_1)$.

◆ The gain-bandwidth product is a measure of amplifier capability.

$$\text{FET: } A_O \omega_2 \le g_m/C_{gs} \qquad \text{BJT: } A_O f_2 \le \beta f_\beta = f_T$$

◆ The benefits of negative feedback outweigh the loss in gain:

Gain can be made insensitive to device parameters.
Operating point can be made insensitive to changes.
Gain can be made insensitive to frequency.
Noise and distortion can be discriminated against.
Input and output impedances can be improved.

- An oscillator is a self-excited amplifier in which a portion of the output signal is fed back in the correct phase and amplitude to make **GH** $= 1 + j0$. By making **G** or **H** frequency dependent, oscillation at a single frequency can be obtained.

- In emitter (source) followers, the entire output is fed back into the input. Gain is less then unity but input and output impedances are optimized. The emitter follower acts as an impedance transformer with a ratio of approximately $\beta + 1$. The source follower offers similar advantages.

REVIEW QUESTIONS

1. What determines whether a signal is "small" or "large"?
2. Distinguish between midfrequency range and bandwidth.
3. In Fig. 18.3, explain the effects of eliminating C_S and halving C_{C1}.
4. Draw a wiring diagram of one-stage of an *RC*-coupled amplifier using an E MOSFET and draw a linear circuit model that is valid for frequencies from 20 Hz to 200 kHz.
5. For the circuit model of the preceding question, state reasonable assumptions and derive circuit models valid for low, medium, and high frequencies.
6. Define in words and in symbols the cutoff frequencies.
7. Differentiate between coupling and bypass capacitors.
8. What limits the low-frequency response of an FET? Of a BJT?
9. What limits the high-frequency response of an FET? Of a BJT?
10. Why is emphasis placed on voltage gain of an FET amplifier and current gain of a BJT amplifier?
11. Explain the Miller technique for transforming feedback circuits. Why is it useful?
12. Compare FET and BJT amplifiers on the following characteristics: input impedance, voltage gain, current gain, and linearity.
13. What is meant by "cascading"? Where is it used?
14. What are the advantages of the decibel as a unit?
15. Define the "gain-bandwidth product." Why is it useful?
16. Is oscillation possible in an ordinary two-stage *RC*-coupled audio amplifier? Explain.
17. A portable high-gain amplifier tends to become oscillatory as the batteries age. It is known that the internal resistance of batteries increases with age. Explain the connection between these phenomena.
18. Explain the expression "trading gain for bandwidth."
19. Draw a block diagram to represent a "noisy" amplifier in which noise (perhaps an objectionable hum) appears in the output without a corresponding input. Suggest a means for reducing such noise.
20. Explain how an amplifier can be made insensitive to the effect of increased temperatures on transistor parameters.
21. Define the impedance characteristics of an ideal amplifier and explain your answer.
22. Why is a common-drain amplifier called a source follower?
23. Cite two virtues of the source follower.
24. Explain how an emitter follower functions as an impedance transformer. Where in a "hi-fi" set would this property be valuable?

EXERCISES

1. A silicon transistor ($\beta = 70$, $C_\pi = 20$ pF, $C_{jc} = 5$ pF, and $r_\pi = 2.8$ kΩ) is used in the circuit of Fig. 18.3b with $R_C = 10$ kΩ, $C_{C1} = C_{C2} = 5\ \mu$F, $R_E = 2$ kΩ, $C_E = 50\ \mu$F, $R_1 = 10$ kΩ, and $R_2 = 100$ kΩ. It is estimated that $C_L + C_W = 50$ pF.

List the assumptions made in deriving the circuit of Fig. 18.5b and check their validity at $f = 5000$ Hz.

2. In the generalized amplifier of Fig. 18.2, $\mathbf{Z}_s \cong R_s = 2\ \text{k}\Omega$, $\mathbf{Z}_i$ consists of $R_i = 1\ \text{M}\Omega$ in parallel with $C_i = 100$ pF, $\mathbf{G} = 0.01\underline{/180^\circ}$, and $\mathbf{Z}_o \cong R_o = 10\ \text{k}\Omega$.
   (a) Assuming $\mathbf{Z}_L$ is a pure resistance R_L, specify R_L for a voltage gain $|\mathbf{V}_o/\mathbf{V}_s| = 50$ at $\omega = 10$ krad/s.
   (b) Predict the voltage gain (magnitude and phase angle) at $\omega = 6.67$ Mrad/s.
3. A BJT amplifier has a midfrequency gain of 200 with upper and lower cutoff frequencies at 10 Hz and 100 kHz, respectively. Draw the frequency-response curve showing gain and phase angle vs. frequency. What is the bandwidth of this amplifier?
4. If $r_\pi \cong R_s$ and $R_B \cong 10r_\pi$ in Fig. 18.5b calculate the error made in predicting A_V by neglecting R_B. Draw a conclusion.
5. An FET for which $g_m = 2\ \text{m}\mho$, $r_d = 200\ \text{k}\Omega$, $C_{gs} = 10$ pF, and $C_{gd} \cong 0$ is used in an RC-coupled amplifier with $R_s = 1\ \text{k}\Omega$, $R_G = 0.1\ \text{M}\Omega$, $R_D = 5\ \text{k}\Omega$, $R_S = 2\ \text{k}\Omega$, and R_L is "very large," $C_{C1} = C_{C2} = 0.1\ \mu\text{F}$, and C_S is "large"; $C_W + C_L$ is estimated to be 10 pF.
   (a) Draw and label a linear circuit model valid for a wide range of frequencies, stating any useful assumptions.
   (b) Draw and label a simplified circuit model for medium frequencies, and predict the midfrequency gain.
6. An enhancement MOSFET ($g_m = 5\ \text{m}\mho$) with voltage-divider bias (Fig. 16.8b) is used as an RC-coupled amplifier.
   (a) Draw a circuit diagram similar to Fig. 18.29, with $R_1 = R_2 = 50\ \text{M}\Omega$, $R_S = 2.5\ \text{k}\Omega$, and $V_{DD} = 30$ V.
   (b) Draw a circuit model valid for moderate frequencies, and specify R_D for a voltage gain $|\mathbf{V}_o/\mathbf{V}_i| = 40$.
   (c) For the specified R_D, $R_s = 1\ \text{k}\Omega$, and $R_L = 20\ \text{k}\Omega$ connected across $\mathbf{V}_o$, predict $\mathbf{V}_o/\mathbf{V}_s$.

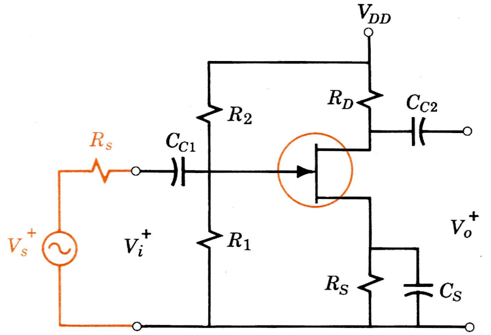


Figure 18.29

7. In the circuit of Fig. 18.30, $R_1 = 20\ \text{k}\Omega$, $R_2 = 80\ \text{k}\Omega$, $R_E = 2\ \text{k}\Omega$, and $V_{CC} = 10$ V. For this silicon transistor, $\beta = 100$.

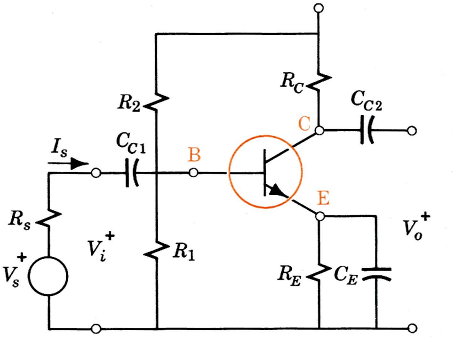


Figure 18.30

   (a) Estimate I_C (by neglecting $I_B R_B$) and calculate r_π.
   (b) Specify R_C for a midfrequency voltage gain $|\mathbf{V}_o/\mathbf{V}_i| = 150$.
   (c) For the specified R_C and $R_s = 1\ \text{k}\Omega$, predict $\mathbf{V}_o/\mathbf{V}_s$.
   (d) For the specified R_C and R_s, predict $\mathbf{V}_o/\mathbf{V}_s$ with $R_L = 10\ \text{k}\Omega$ connected across $\mathbf{V}_o$.
8. In the circuit of Fig. 18.30, $R_1 = 10\ \text{k}\Omega$, $R_2 = 90\ \text{k}\Omega$, $R_C = 10\ \text{k}\Omega$, and $I_C = 1.25$ mA. For this transistor $\beta = 50$.
   (a) Replace the transistor by an appropriate linear model and redraw the circuit to permit the prediction of the small-signal performance at moderate frequencies. (Show terminals B, C, and E on your diagram.)
   (b) Predict the midfrequency current gain I_{b2}/I_{b1} if the following stage employs a similar transistor under similar conditions.

(c) If $R_s = 1\,\text{k}\Omega$, predict the overall voltage gain V_2/V_s.

9. In the circuit of Fig. 18.30, $R_1 = 20\,\text{k}\Omega$, $R_2 = 60\,\text{k}\Omega$, $I_C = 2\,\text{mA}$, and $\beta = 80$. The source is a phono pickup represented by $V_s = 10\,\text{mV}$ (rms) in series with $R_s = 500\,\Omega$. The "load" is another stage whose input resistance is $10\,\text{k}\Omega$. Complete the design so that V_o is 1.6 V (rms) at midfrequencies.
   (a) Draw an appropriate small-signal model.
   (b) Specify the value of R_C.
   (c) Specify values of R_E and V_{CC}.
10. For a certain RC-coupled amplifier, $\omega_{11} = \omega_{12} = 100\,\text{Hz}$. What is ω_1 in this case?
11. A single-stage FET RC coupled amplifier used in a mechanical engineering experiment must have a gain at 1 Hz equal to 90% of its midfrequency gain.
   (a) What lower cutoff frequency is required?
   (b) If $g_m = 2\,\text{m}\mho$, $R_D = 10\,\text{k}\Omega$, and if the input resistance to the following stage is $1\,\text{M}\Omega$, what value of C_{C2} is required?
   (c) If the gain is halved by reducing R_D, is the required value of C_{C2} reduced?
12. In the amplifier of Exercise 8, $R_E = 1\,\text{k}\Omega$, $C_E = 50\,\mu\text{F}$, and $C_{C1} = 5\,\mu\text{F}$. C_{C2} is "large."
   (a) Calculate the "lower cutoff frequencies" as defined by C_E and C_{C1} separately.
   (b) What is the lower cutoff frequency for the amplifier?
   (c) Specify an appropriate value for C_{C2}, explaining your reasoning.
13. In the amplifier of Exercise 8, $R_E = 1\,\text{k}\Omega$.
   (a) Specify the value of C_E for a lower cutoff frequency of 50 Hz.
   (b) Specify appropriate values of C_{C1} and C_{C2}.
14. For a certain n-channel JFET, the manufacturer specifies $g_m = 4000\,\mu\mho$, and maximum values of $C_{iss} = 7\,\text{pF}$ and $C_{rss} = 3\,\text{pF}$. This JFET is used in the amplifier of Fig. 18.3a with $R_s = 1\,\text{k}\Omega$, $R_1 = 1\,\text{M}\Omega$, $R_2 = 1.5\,\text{M}\Omega$, $C_{C1} = C_{C2} = 0.1\,\mu\text{F}$, $R_D = 5\,\text{k}\Omega$, $R_S = 10\,\text{k}\Omega$, $C_S = 50\,\mu\text{F}$, and $R_L = 1\,\text{M}\Omega$.
   (a) Draw an appropriate small-signal model and predict the midfrequency gain V_o/V_s.
   (b) Draw an appropriate small-signal model and predict the lower cutoff frequency.
15. For a certain silicon transistor in the common-emitter configuration, the small-signal current relations are:

$$\begin{cases} i_b = 0.004v_{be} + 10^{-10}\,dv_{be}/dt \\ i_c = 0.02v_{be} + 2 \times 10^{-5}v_{ce} \end{cases}$$

   (a) Draw and label an appropriate circuit model.
   (b) What determines the frequency response of this device?
   (c) Determine the frequency at which the base signal voltage (and therefore the output current) is only 70.7% of its low-frequency value for a given input current.
16. The enhancement MOSFET of Fig. 13.6 is used as an amplifier with V_s in series with R_s applied from gate to ground and V_o taken between drain and ground.
   (a) Replace the MOSFET by a small-signal model and use nodal analysis to evaluate $A = |V_o/V_s|$ for $R_s = 10\,\text{k}\Omega$, $R_G = 1\,\text{M}\Omega$, $g_m = 2\,\text{m}\mho$, and $R_D = 10\,\text{k}\Omega$.
   (b) Derive an expression for input resistance of the amplifier and evaluate R_i.
   (c) Replace R_G by its Miller equivalent, repeat parts (a) and (b), and draw a conclusion.
17. For the FET amplifier of Exercise 5,
   (a) Draw and label a simplified circuit model valid for low frequencies and predict ω_1 and f_1.
   (b) Draw and label a simplified circuit model valid for high frequencies and predict ω_2 and f_2.
   (c) For this amplifier, define "midfrequency" and sketch the frequency response curve. (Plot 3 key points on the A_V vs f curve; for a log plot on ordinary graph paper let unit distance correspond to a factor of 10 in frequency.)
18. For a certain RC-coupled amplifier, $\omega_{21} = \omega_{22} = 100\,\text{kHz}$. What is ω_2 in this case?
19. For the JFET amplifier of Exercise 14, draw an appropriate small-signal model

and predict the upper cutoff frequency, assuming about 4 pF of wiring capacitance in the drain circuit.

20. For the 2N699B (see Table 17-1), beta is down to unity at 70 MHz.
   (a) Devise an appropriate small-signal model and evaluate the parameters.
   (b) Estimate the upper cutoff frequency of the current gain in an amplifier using the 2N699B, stating any assumptions made.

21. In the amplifier of Exercise 20, $R_B = R_1 \| R_2$ is negligibly large, wiring capacitance is negligibly small, $C_{jc} = 13$ pF, and $R_o = 2\ \text{k}\Omega$. Predict the midfrequency voltage gain and the upper cutoff frequency.

22. Explain why a Bode plot of A where $\mathbf{A} = 1/(1 + j\omega RC)$ above $\omega_2 = 1/RC$ is a straight line of slope -20 dB per decade.

23. A single-stage transistor amplifier has a midfrequency current gain of 40 and a lower cutoff frequency of 60 Hz. Predict the corresponding values for an amplifier consisting of three such stages in cascade.

24. The output of a single-stage of transistor amplification is 2 V (rms) across a load resistance of 10 kΩ. The input is 0.5 mV (rms) across an input resistance of 1 kΩ.
   (a) Express the power gain in decibels
   (b) Express the voltage gain and current gain in decibels.

25. The power loss in 1 kilometer of coaxial telephone cable is 1.5 dB.
   (a) What is the power loss in 2 km of telephone cable? Three?
   (b) What is the total power loss from New York to Los Angeles?
   (c) What signal power input is required at New York to deliver 1 μW at Los Angeles? Would amplifiers be desirable?

26. A telephone system using cable with a power loss of 1.5 dB/kilometer requires a minimum signal of 1 μW.
   (a) If amplifiers are located 50 km apart, what amplifier output power is required?
   (b) Assuming 10% overall amplifier efficiency, what total (dc) input power is required for New York to Los Angeles transmission?
   (c) Compare the result of part (b) to the answer for part (c) of Exercise 25.

27. An amplifier consists of the following stages: Stage 1, midfrequency power gain = 10 dB, $f_1 = 0$ Hz, $f_2 = 3$ MHz. Stage 2, midfrequency power gain = 30 dB, $f_1 = 10$ Hz, $f_2 = 1$ MHz. Stage 3, midfrequency power gain = 20 dB, $f_1 = 100$ Hz, $f_2 = 300$ kHz.
   (a) Plot the individual and overall response curves.
   (b) If the midfrequency output is to be 2 W, what input power is required?

28. A transistor with a nominal $\beta = 50$ has a gain-bandwidth product $f_T = 30$ MHz. Describe, quantitatively, the high-frequency characteristics of this transistor. What is the current gain in dB at 30 MHz? At 2 MHz?

29. Each stage of a transistor amplifier is to provide a current gain of approximately 20 with $f_2 = 5$ MHz. For a $\beta = 80$, what minimum beta cutoff frequency is required?

30. A multistage amplifier is to provide a voltage gain of 1000 with an upper cutoff frequency of 200 kHz. Available is an FET for which $g_m = 2\ \text{m}\mho$, $C_{rss} = 2$ pF, and $C_{iss} = 30$ pF. How many stages are required?

31. For the circuit of Fig. 18.31, draw an ac circuit model.

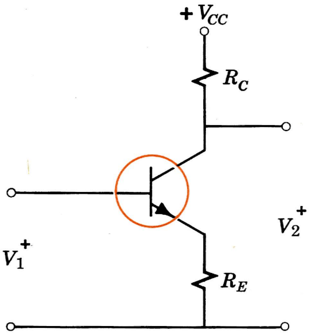


Figure 18.31

   (a) Note the polarities of ac voltages and determine the voltage feedback factor H.
   (b) Draw a block diagram for this circuit.

(c) Derive an expression for overall midfrequency gain from the circuit model, and from the block diagram.

32. Repeat Exercise 31 for the circuit of Fig. 18.32, identifying *current* feedback.

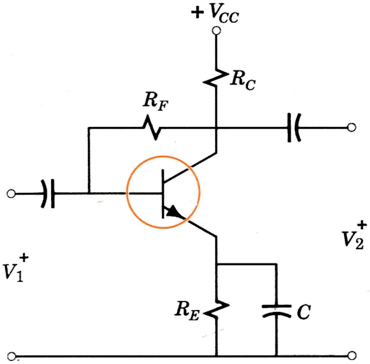


Figure 18.32

33. A feedback system is represented by Fig. 18.19a with $\mathbf{H} = 0.005\,\underline{/0^\circ}$. Define the conditions necessary for oscillation.

34. For the circuit of Fig. 18.19b,
(a) Determine the frequency of oscillation in words and symbols.
(b) If $C_1 = C_2 = 200$ pF and $L = 25$ mH, estimate the oscillation frequency.

35. In the phase-shift oscillator of Fig. 18.20, $R = 100$ kΩ and $R_D = 10$ kΩ.
(a) Determine the approximate value of C for oscillation at 20 Hz; state any simplifying assumptions made.
(b) If oscillation at 20 Hz is to be determined by a resonant circuit with $L = 1$H, approximately what value of C is required? Is this a reasonable value for a variable capacitor?

36. For frequencies below $\omega = 2000$ rad/s, the gain of an amplifier is described by $\mathbf{G} = 0.1\omega\,\underline{/180^\circ}$. For oscillation at $\omega = 1500$ rad/s, what value of $\mathbf{H}$ is required?

37. A common-emitter small-signal amplifier (Fig. 18.5b) uses a 2N699B with $I_C = 1$ mA and $V_{CE} = 5$ V. $R_C = 3$ kΩ and $R_B = 30$ kΩ.
(a) Through what range can the *per stage* current gain vary without departing from the manufacturer's specifications (min to max)?
(b) Introduce current feedback with $H = -0.2$ and recalculate the range of current gain.
(c) How could the loss in current gain due to the introduction of negative feedback be made up?

38. As the ambient temperature increases, input resistance h_{ie} increases more rapidly than $h_{fe}(\beta)$. As a result, for a certain transistor amplifier with a voltage gain of 200, a temperature increase of 50°C results in a 10% decrease in gain. You are to use feedback to reduce the net effect on overall gain to 2%.
(a) Draw a block diagram of the amplifier, labeling V_1 and V_2.
(b) Add feedback and specify the necessary feedback factor.

39. The high-frequency response of an amplifier is defined by $A = A_0/(1 + j\omega/\omega_2)$ where $A_0 = 100$ and $\omega_2 = 10$ krad/s.
(a) Plot a couple of points and sketch a graph of A/A_0 versus ω on a logarithmic scale.
(b) Assume that a feedback loop with $H = -0.04$ is provided and repeat part (a). Label this graph "With feedback."

40. An amplifier has a midfrequency gain of 100, but the gain is only 50 at 10 Hz. Devise a feedback system and specify a real value of H so that the gain at 10 Hz is within 5% of the new midfrequency gain.

41. In Fig. 18.25a, consider V_D to be the input and V_o the output and redraw the block diagram. Calculate the gain V_o/V_D for this system and interpret the result.

42. An audiofrequency amplifier with a voltage gain of 200 has an output of 5 V, but superimposed on the output is an annoying 60-Hz hum of magnitude 0.5 V.
(a) Devise a feedback system and specify H so that with the same 5-V output, the hum component is reduced to 0.025 V.
(b) Draw separate block diagrams for the audio signal and the hum component, showing the actual voltages at each point in the system.

43. A class-A power amplifier (see Fig. 16.21a) using a 2N3114 is to supply 70 V (rms) of signal across a 10,000-Ω load resistance.

(a) How much ac power is supplied? To get this much power, the transistor must be driven into the nonlinear regions. It is estimated that the major element of distortion is a 7-V (rms) second harmonic. Express this as percent distortion.

(b) The source for the amplifier is a phono pickup with an output of 6 mV and a source resistance of 1500 Ω. How many amplifier stages (operating at $I_C = 1$ mA) of what current gain will be required to drive the final amplifier?

(c) It is suggested that the distortion can be reduced by providing feedback around the power amplifier. Specify the feedback factor necessary to reduce the distortion to 1%.

(d) To maintain the required output voltage, how must the intermediate amplifier be modified?

44. An emitter follower employs a 2N699B at $I_C = 1$ mA supplied by a source for which $V_s = 2$ V and $R_s = 2$ kΩ. For an emitter resistance $R_L = 1$ kΩ, predict the output voltage and the input and output resistances.

45. A device with a high output resistance ($R_s = 50$ kΩ) is to be coupled to a load resistance R_L by means of a 2N3114 transistor.

(a) Devise a circuit and specify R_L so that the input resistance seen by the source is approximately 50 kΩ.

(b) Predict the output resistance of the transistor under these circumstances.

46. The input to a device can be represented by $R = 100$ kΩ in parallel with $C = 100$ pF. It is fed by an FET for which $g_m = 2$ m℧. Predict the upper cut-off frequency if the FET is used as:

(a) an ordinary amplifier with $R_D = 5$ kΩ.

(b) a source follower with $R_S = 5$ kΩ.

47. Modify the circuit of Fig. 18.28a so that V_i is applied (through C_C) between the gate and source terminals. Draw the small-signal model and express V_o/V_i in terms of the circuit parameters. Use this amplifier as one block in a block diagram representation of Eq. 18-60. What is the feedback factor H?

PROBLEMS

1. An inexperienced designer omits the source by-pass capacitor C_S in Fig. 18.29. As a first step in determining the effect of the omission at midfrequencies, replace the FET by a small-signal model and redraw the circuit. Derive an expression for V_o/V_i and evaluate the effect on voltage gain if $R_S = 5(1/g_m)$. What advantage would such a circuit offer?

2. Derive the criterion for C_E (Eq. 18-21) by using the approach followed in deriving the criterion for C_S (Eq. 18-17).

(a) Define the condition for ω_1.

(b) Determine the Thévenin equivalent resistance across which C_E is connected (Fig. 18.11).

(c) Express ω_1 in terms of the circuit parameters.

3. An FET amplifier similar to Fig. 18.3a with $R_D = 5$ kΩ is built and tested. The manufacturer specifies $g_m = 4000$ μ℧, $C_{rss} = 5$ pF and $C_{iss} = 25$ pF for the FET. When V_o/V_i is observed on a CRO with an input capacitance of 30 pF, the midfrequency voltage gain is observed to be 18 and the gain is down to 12.7 at 1 MHz. Analyze these experimental results and draw a conclusion.

4. A small-signal wideband amplifier is to provide a voltage gain of 40 with a flat response from 20 Hz to approximately 100 kHz. Available is a 2N699B transistor (see Table 17-1, p. 458) with $\beta_f = 3.5$ at 20 MHz and $C_{jc} = 13$ pF. The resistance of the source is $R_s = 700$ Ω. Stating *all* assumptions, design the amplifier (draw the wiring diagram and specify the components and supply voltage).

5. The "Darlington pair" shown in Fig. 18.33 is frequently packaged as a single component or made available as an integrated circuit. Assuming identical transistors

(β, r_π) and stating any other necessary assumptions, predict the current gain of the pair and the input and output resistances of the pair in the given circuit. Draw a conclusion about the advantages of such a component.

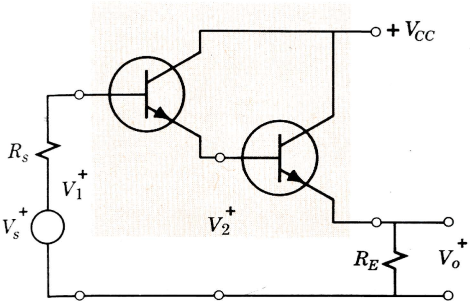


Figure 18.33

6. Ten watts of audio power is to be supplied to a 250-Ω load.
   (a) What output voltage (rms) and current are required? For this voltage swing, the transistor must be driven into nonlinear regions resulting in distortion in the form of a 5-V (rms) second harmonic. Express this as "% distortion."
   (b) The available phono pickup has an output of 1 mV. If the final power amplifier has a voltage gain of 50, specify the number and gain of the necessary voltage amplifiers.
   (c) Show how distortion can be reduced by providing feedback around the final amplifier and specify the feedback factor to reduce the distortion to 1%. Specify the changes necessary in the voltage amplifiers to maintain the same output voltage.
7. An amplifier is represented by the general model of Fig. 18.2 with *R*s in place of **Z**s. Reproduce the circuit and add negative voltage feedback with a feedback factor *H*. Derive expressions for G_F, R_{oF}, and R_{iF}, the new gain, output resistance, and input resistance. Draw a general conclusion.
8. An innovative experimenter disconnects C_E in Fig. 18.30 and finds the midfrequency voltage gain V_o/V_i greatly reduced. He expects that there may be some offsetting improvement and asks you to analyze the performance. Derive an expression for V_o/V_i and predict the effect of disconnecting C_E if $\beta = 100$, $R_C = 5\ \text{k}\Omega$, $R_E = 1\ \text{k}\Omega$, and $I_C = 1$ mA. Investigate the new input impedance and draw a conclusion.
9. A JFET for which $I_{DSS} = 5$ mA, $V_{p.} = -3$ V, $C_{gs} = 20$ pF, and $C_{gd} = 2$ pF is used as a source-follower in the circuit of Fig. 18.28a with $R_G = 5\ \text{M}\Omega$ and $V_{DD} = 20$ V. For quiescent operation at $I_D = 2$ mA and $V_{DS} = 10$ V, design the circuit, i.e., specify R_1 and R_2. Predict the voltage gain, the output resistance, and the input impedance at 10 kHz.
10. The "hybrid" circuit shown in Fig. 18.34 provides a very high input impedance and a very low output impedance.
    (a) Describe the operation of the circuit in a qualitative way.
    (b) Replace the active devices by simple small-signal models and redraw the circuit for convenient analysis.
    (c) For $R_1 = 10\ \text{M}\Omega$, $g_{m1} = 3 \times 10^{-3}$ ℧, $I_C = 1$ mA, $\beta_2 = 100$, and $R_3 = 3\ \text{k}\Omega$, predict the voltage gain V_3/V_1, R_i, and R_o.

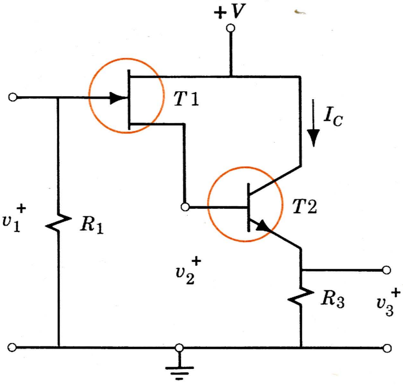


Figure 18.34

Chapter 19

Operational Amplifiers

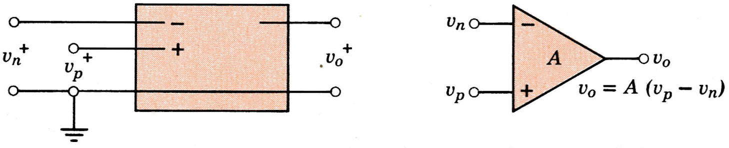
Ideal Op Amps
Typical Applications
Practical Considerations
The Analog Computer

The name *operational amplifier* was first applied to the electronic amplifiers employed in analog computers to perform mathematical operations such as summing and integration. The early vacuum-tube designs were complex, bulky, and expensive. Today, sophisticated integrated-circuit operational amplifiers can be fabricated on a single chip of silicon and sold for less than a dollar. As a result, the design of electronic signal processing equipment has been radically altered. For linear or analog systems, the IC op amp plays the same basic-building-block role as do the IC gates and counters in digital systems.

Our purpose here is to learn how these versatile amplifiers work and how they can be used in practical circuits. We start by considering ideal operational amplifiers and look at some basic circuits and general applications. Then we consider some practical limitations and design aspects. Finally we examine the electronic analog computer to see how operational amplifiers can be arranged to simulate physical systems.

Ideal Operational Amplifiers

An *op amp* is a direct-coupled, high-gain voltage amplifier designed to amplify signals over a wide frequency range. Typically, it has two input terminals and one output terminal and a gain of at least 10^5 and is represented by the symbol in Fig. 19.1b. It is basically a *differential amplifier* responding to the difference



(a) Amplifier diagram (b) Op amp symbol

Figure 19.1 Operational amplifier diagram and symbol.

in the voltages applied to the positive and negative input terminals. (Single-input op amps correspond to the special case where the + input is grounded.) It is normally used with external feedback networks that determine the function performed.

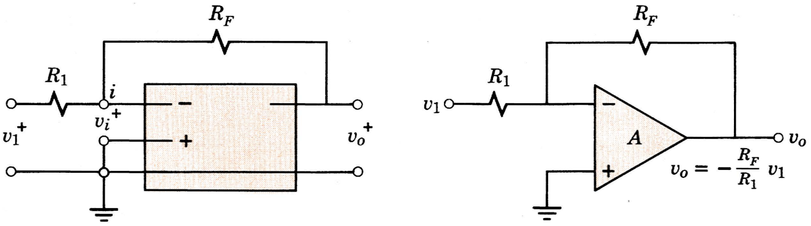
The characteristics of the ideal op amp are as follows:

Voltage gain $A = \infty$
Output voltage $v_o = 0$ when $v_n = v_p$
Bandwidth $BW = \infty$
Input impedance $Z_i = \infty$
Output impedance $Z_o = 0$

Although these are extreme specifications, commercially available units approach the ideal so closely that many practical circuits can be designed on the basis of these characteristics. The design problem is reduced to selecting the commercial unit that is nearly ideal in the parameters that are significant in a given situation.

Basic Inverting Circuit[†]

In the circuit of Fig. 19.2a, the input signal is applied to the negative or *inverting* terminal and the positive or *noninverting* terminal is grounded. The input voltage v_1 is applied in series with input resistance R_1, and the feedback resistance R_F is connected between output and input. Because of the high input impedance to the amplifier, the amplifier input current is negligibly small. Therefore, the input current is just equal to the feedback current and at node *i* we can write



(a) Amplifier diagram (b) Symbolic representation

Figure 19.2 Basic inverting amplifier circuit.

$$\frac{v_1 - v_i}{R_1} + \frac{v_o - v_i}{R_F} = 0 \tag{19-1}$$

Since the gain of the amplifier is A, the voltage $v_i = -v_o/A$ and Eq. 19-1 can be rewritten as

---

[†] Unfortunately, the term "operational amplifier" is applied to the basic high-gain amplifier as well as to the functional circuit in which it is employed. To lessen confusion, we shall start off by specifying "circuit" when the latter is intended. Later on, we shall use op amp for both, as is done in practice.

$$\frac{v_1}{R_1} + \frac{v_o}{R_1 A} + \frac{v_o}{R_F} + \frac{v_o}{R_F A} = 0$$

Solving for the amplifier output,

$$v_o = \frac{-v_1(R_F/R_1)}{1 + (1/A)(1 + R_F/R_1)} \tag{19-2}$$

Since A is at least 10^5 whereas R_F and R_1 are usually chosen to be of the same order of magnitude, $A \gg (1 + R_F/R_1)$ and the output is, with little error,

$$v_o = -\frac{R_F}{R_1} v_1 \tag{19-3}$$

This relation describes an *inverting amplifier* with gain $A_F = -R_F/R_1$. (The same relation can be derived using the feedback approach or the Miller technique, see Problems 1 and 2.) From this analysis, we can draw the following conclusions:

1. Since $v_i = -v_o/A$ is so small compared to v_o and v_1, input current $i_1 = v_1/R_1$ and is independent of R_F. Therefore, the input is isolated from the output.
2. The gain of the amplifier circuit is independent of the voltage gain of the amplifier as long as it is "high." Therefore, gain A_F will be unaffected by changes in temperature, device parameters, or frequency.

The first conclusion leads to a simple circuit model for the inverting amplifier (Fig. 19.3). The second indicates an easy way to design an amplifier.

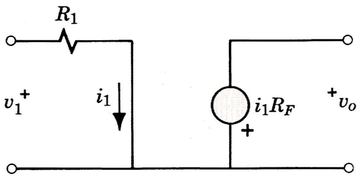


Figure 19.3 Circuit model for an inverting amplifier.

Example 1

Design an amplifier with a voltage gain of −100, an input resistance of 1000 Ω, and a flat response over the audio-frequency range.

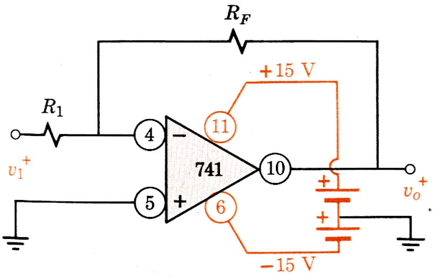


Figure 19.4 Amplifier design.

In this case, "design" means selecting an appropriate op amp, specifying the feedback network, and providing the necessary power supply.

The type 741 (see Appendix 4) is a high performance op amp suitable for a variety of analog applications at audio frequencies. It has dual-inline pin connections as shown in Fig. 19.4 and operates from ±15-V power supplies or batteries.

The network design proceeds as follows:

$$R_1 = R_{iF} = 1000\ \Omega$$

$$R_F = -R_1 \times A_F = R_1 \times 100 = 100\ \text{k}\Omega$$

For noncritical applications, this is an adequate circuit.

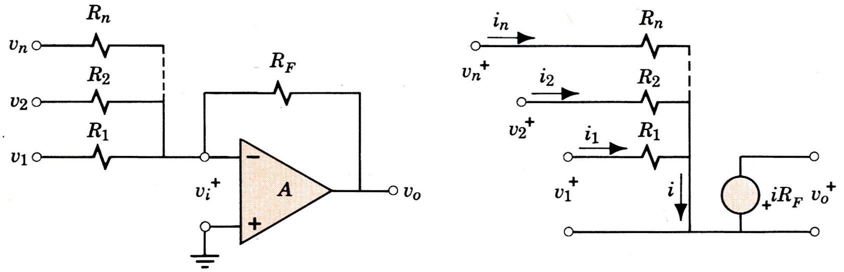


Figure 19.5 Summing circuit; $v_o = -[v_1(R_F/R_1) + v_2(R_F/R_2) + \cdots + v_n(R_F/R_n)]$.

Summing Circuit

If the circuit of Fig. 19.2 is modified to permit multiple inputs, the operational amplifier can perform addition. In Fig. 19.5, the input current is supplied by several voltages through separate resistances. If the input impedance of the amplifier is very high, the sum of the input currents is just equal and opposite to the feedback current or

$$\frac{v_1 - v_i}{R_1} + \frac{v_2 - v_i}{R_2} + \cdots + \frac{v_n - v_i}{R_n} = \frac{v_i - v_o}{R_F} \tag{19-4}$$

Again the gain A is very large and since $v_i = v_o/A$, v_i is very small compared to v_o, v_1, v_2, etc. Neglecting v_i as before Eq. 19-4 becomes

$$\frac{v_1}{R_1} + \frac{v_2}{R_2} + \cdots + \frac{v_n}{R_n} = -\frac{v_o}{R_F}$$

or

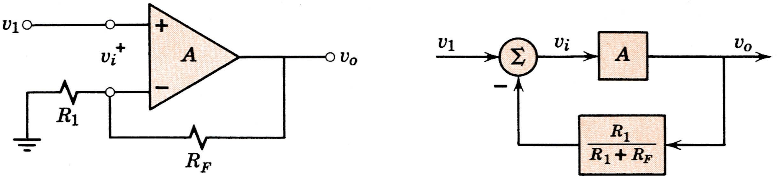
$$v_o = -\left(\frac{R_F}{R_1}v_1 + \frac{R_F}{R_2}v_2 + \cdots + \frac{R_F}{R_n}v_n\right) \tag{19-5}$$

As would be expected from looking at the circuit model, the output of the amplifier circuit is the weighted sum of the inputs. This is an analog *adder* in contrast to the digital adder shown in Fig. 15.7. Note that a signal can be subtracted by first passing it through an analog *inverter* consisting of an inverting amplifier with $R_F = R_1$.

Basic Noninverting Circuit

If the basic amplifier is reconnected as in Fig. 19.6a, the behavior is modified. Here the signal is applied to the noninverting + terminal and a fraction of the output signal is fed back to the − terminal. The governing equations are:

$$v_o = Av_i \qquad \text{and} \qquad v_i = v_1 - \frac{R_1}{R_1 + R_F}v_o \tag{19-6}$$



(a) Symbolic representation (b) Block diagram

Figure 19.6 The noninverting amplifier circuit.

These relations are realized in the block diagram of Fig. 19.6b. Instead of an algebraic solution, we can write by inspection

$$A_F = \frac{G}{1 + GH} = \frac{A}{1 + AR_1/(R_1 + R_F)} \cong \frac{R_1 + R_F}{R_1} \tag{19-7}$$

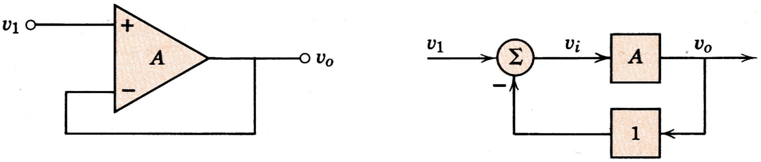
if A is large. Here again the gain is determined by the feedback network elements R_1 and R_F. For the noninverting circuit, however, the gain is positive and is equal to or greater than unity. Furthermore we expect a very large input resistance $R_{in} = v_{in}/i_{in} = v_1/i_i$. Any increase in v_1 which would tend to raise v_i and i_i is offset by high-gain feedback from v_o that tends to prevent any change in v_i and i_i.

Voltage Follower

A useful special case of the noninverting circuit is shown in Fig. 19.7a. Here $R_F = 0$ and $R_1 = \infty$ (open circuit). From Eq. 19.7, the circuit gain is now

$$A_F = \frac{v_o}{v_1} \cong \frac{R_1 + R_F}{R_1} = 1 \tag{19-8}$$

Alternatively, A_F can be calculated from the block diagram of Fig. 19.7b with the same result. In either case, the output voltage is just equal to the input voltage and this is a *voltage follower*, a more refined version of the emitter follower or source follower.

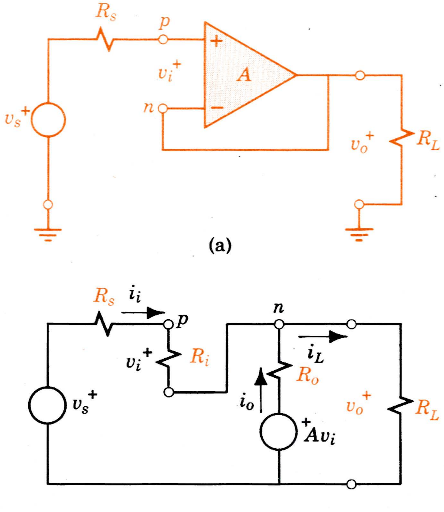


(a) Symbolic representation (b) Block diagram

Figure 19.7 The voltage-follower circuit.

Example 2

In the circuit of Fig. 19.8a, $R_s = 1\ \text{k}\Omega$ and $R_L = 10\ \text{k}\Omega$. For the op amp, $A = 10^5$, $R_i = 100\ \text{k}\Omega$ and $R_o = 100\ \Omega$. For $v_o = 10$ V, calculate v_s and v_o/v_s and estimate the input resistance of the circuit.



(a)

(b)

Figure 19.8 Voltage follower.

The circuit is redrawn in Fig. 19.8b to show R_i in series with the input and R_o in series with Av_i. For $v_o = 10$ V,

$$i_L = \frac{v_o}{R_L} = \frac{10}{10^4} = 10^{-3}\ \text{A}$$

Expecting i_i to be very small, we write

$$\begin{aligned} Av_i &= v_o + i_oR_o \cong v_o + i_LR_o \\ &= 10 + 10^{-3} \times 10^2 = 10.1\ \text{V} \\ \therefore v_i &= (Av_i)/A = 10.1 \times 10^{-5}\ \text{V} \end{aligned}$$

Hence $i_i = v_i/R_i = v_i/10^5 = 1.01 \times 10^{-9}$ A and the assumption regarding i_i is justified.

$$\begin{aligned} v_s &= v_o + i_i(R_s + R_i) \\ &= 10 + 1.01 \times 10^{-9}(1.01 \times 10^5) \\ &= 10.0001\ \text{V} \\ v_o/v_s &= 10/10.0001 = 0.99999 \end{aligned}$$

This is indeed a voltage follower with unity gain.

With feedback, the input resistance is

$$R_{iF} \cong \frac{v_s}{i_i} \cong \frac{10}{1.01 \times 10^{-9}} \cong 10^{10}\ \Omega$$

a very high value.

Typical Applications

The approach in operational amplifier circuit design is to use a very high-gain unit with feedback to provide a lower but more stable effective gain, improved impedance characteristics, and better frequency response. In the *inverting amplifier,* gain $A_F = -R_F/R_1$, input resistance $R_{iF} = R_1$, and the bandwidth extends from zero to $f_{2F} = f_2(1 + GH) = f_T/A_F$, where f_T is the unity-gain frequency of the op amp unit. (See Eqs. 18-44 and 18-48.)

The *noninverting amplifier* performs similarly to the inverting type, but it has two distinctive characteristics. First, output signals are in phase with those at the input. Second, the input resistance is very high, approaching infinity in practical terms, and the output resistance is very low. This means that noninverting amplifiers do not "load" their sources and, in turn, they are not affected by other loads.

Differential Amplifier

The *direct-coupled differential amplifier* shown in Fig. 19.9 is widely used in instrumentation. In the typical transducer, a physical variable such as a change in temperature or pressure is converted to an electrical signal that is then amplified to a useful level. The differential amplifier is advantageous because it discriminates against dc variations, drifts, and noise and responds only to the significant changes.

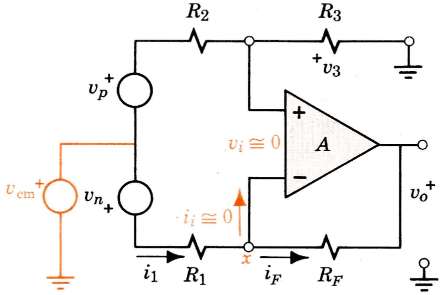


Figure 19.9 A direct-coupled differential amplifier.

In Fig. 19.9, a transducer output is represented by a *common-mode voltage* v_{cm} and a differential voltage $v_p - v_n$. Assuming an ideal op amp, $v_i = 0$ and the potential at node x is equal to the voltage across R_3 or

$$v_3 = (v_{cm} + v_p)\frac{R_3}{R_2 + R_3} = v_x \tag{19-9}$$

Since $i_i = 0$, the currents i_1 and i_F are equal or

$$i_1 = \frac{v_{cm} + v_n - v_x}{R_1} = i_F = \frac{v_x - v_o}{R_F} \tag{19-10}$$

Combining these equations and simplifying yields

$$\begin{aligned} v_o = v_{cm}&\left(\frac{R_3}{R_2 + R_3} - \frac{R_F}{R_1} + \frac{R_F}{R_1}\cdot\frac{R_3}{R_2 + R_3}\right) \\ &+ v_p\left(\frac{R_3}{R_2 + R_3} + \frac{R_F}{R_1}\cdot\frac{R_3}{R_2 + R_3}\right) - v_n\left(\frac{R_F}{R_1}\right) \end{aligned} \tag{19-11}$$

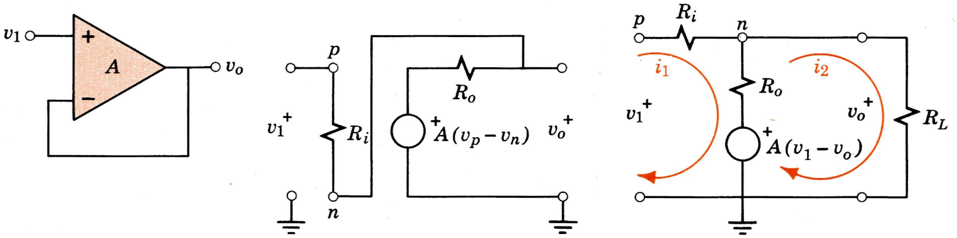
If we design the network so that $R_F/R_1 = R_3/R_2$, the coefficient of v_{cm} goes to zero, the *common-mode signal is rejected*, and the output is

$$v_o = -\frac{R_F}{R_1}(v_n - v_p) \tag{19-12}$$

The ability of a differential amplifier to discriminate against signals that affect both inputs in common and to amplify the small differences in input is a valuable characteristic.

Unity-Gain Buffer

Example 2 demonstrated that the gain of a voltage follower is almost exactly 1; its usefulness lies in its ability to isolate a high-resistance source from a low-resistance load. To provide this isolation, the isolating network should have a very high input resistance and a very low output resistance. We cannot use the ideal op amp model to evaluate the input and output characteristics of such a *unity-gain buffer,* because they depend on the nonideal properties of the operational amplifier. Instead we use the model shown in Fig. 19.10b which includes the finite values of R_i and R_o for the op amp.



(a) Symbol (b) Circuit model (c) Rearranged circuit

Figure 19.10 Analysis of the unity-gain buffer circuit.

Input Resistance. Writing loop equations for the rearranged circuit,

$$v_1 - R_i i_1 - R_o(i_1 - i_2) - Av_1 + Av_o = 0 \tag{19-13}$$

$$v_1 - R_i i_1 - R_L i_2 = 0 \tag{19-14}$$

By inspection, $v_o = R_L i_2$, and from Eq. 19-14, $i_2 = (v_1 - R_i i_1)/R_L$. Substituting these values in Eq. 19-13 yields

$$v_1 - R_i i_1 - R_o i_1 + R_o \frac{v_1 - R_i i_1}{R_L} - Av_1 + AR_L \frac{v_1 - R_i i_1}{R_L} = 0 \tag{19-15}$$

Solving, the input resistance with feedback is

$$R_{iF} = \frac{v_1}{i_1} = \frac{R_i + R_o + (R_o/R_L)R_i + AR_i}{1 + R_o/R_L}$$

$$= \frac{R_L R_i + R_L R_o + R_o R_i + AR_L R_i}{R_L + R_o} \tag{19-16}$$

For the practical case of A very large and $R_L \gg R_o$, this becomes

$$R_{iF} = \frac{(A + 1)R_i R_L}{R_L} + \frac{R_o(R_L + R_i)}{R_L} \cong AR_i \tag{19-17}$$

A buffer using an op amp with $R_i = 100\ \text{k}\Omega$ and $A = 10^5$ would present an input resistance of 10,000 MΩ (as in Example 2).

Output Resistance. Working from Fig. 19.10c, the output resistance can be obtained as v_{OC}/i_{SC}. For $i_2 = 0$,

$$v_{OC} = v_1 - R_i i_1 = v_1 - R_i \frac{v_1 - Av_1 + Av_{OC}}{R_i + R_o} \tag{19-18}$$

Solving,

$$v_{OC} = v_1 \frac{R_i + R_o + (A-1)R_i}{R_i + R_o + AR_i} = v_1 \frac{R_o + AR_i}{R_o + (1+A)R_i} \tag{19-19}$$

For the output shorted, $v_o = 0$ and

$$i_{SC} = \frac{v_1}{R_i} + \frac{Av_1}{R_o} = v_1 \frac{R_o + AR_i}{R_o R_i} \tag{19-20}$$

For the practical case of A very large and $R_i \gg R_o$, these yield

$$R_{oF} = \frac{v_{OC}}{i_{SC}} = \frac{R_o R_i}{R_o + (1+A)R_i} \cong \frac{R_o}{A} \tag{19-21}$$

A buffer using an op amp with $R_o = 100\ \Omega$ and $A = 10^5$ will present an output resistance of $0.001\ \Omega$.

Example 3

An instrumentation transducer is characterized by a voltage $V_T = 5$ V in series with a resistance $R_T = 2000\ \Omega$. It operates an indicator characterized by an input resistance of $100\ \Omega$. Predict the voltage and power delivered to the indicator with and without the use of a buffer.

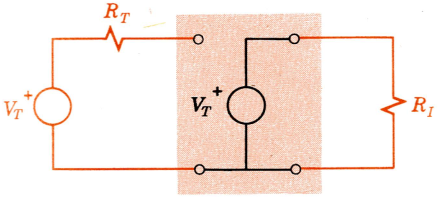


Figure 19.11 Buffer application.

Disregarding the buffer (Fig. 19.11), the voltage across the indicator would be

$$V_I = V_T \frac{R_I}{R_T + R_I} = 5\frac{100}{2100} = 0.238 \text{ V}$$

and the power delivered would be

$$P_I = V_I^2/R_I = (0.238)^2/100 = 0.566 \text{ mW}$$

With a unity-gain buffer, the output voltage follows the input. Since $R_i \to \infty$, no input current flows and the model of the voltage follower is as shown. Now, $V_I = V_T = 5$ V and the power is

$$P_I = V_I^2/R_I = 5^2/100 = 250 \text{ mW}$$

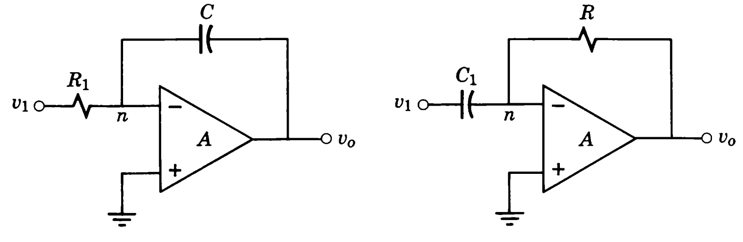
The power gain achieved by inserting the buffer is

$$\frac{P_2}{P_1} = \frac{250}{0.566} = 442$$

Integrator and Differentiator

Up to this point, we have assumed that the feedback network is purely resistive. In general, however, the elements will be complex impedances $\mathbf{Z}_1$ and $\mathbf{Z}_F$ made up of capacitances, inductances, and resistances. Very different functions will be performed as a result of interchanging one pair of network components.

In Fig. 19.12a, the feedback element is a capacitor. For the op amp, $v_i \cong 0$



(a) Integrator (b) Differentiator

Figure 19.12 Electronic integration and differentiation circuits.

and the sum of the currents into node *n* is

$$\frac{v_1}{R_1} + C\frac{dv_o}{dt} = 0 \tag{19-22}$$

Integrating each term with respect to time and solving,

$$v_o = -\frac{1}{R_1 C}\int v_1\, dt \tag{19-23}$$

and the device is an *integrator*. The analog integrator is very useful in computing, signal processing, and signal generating.

If the resistance and capacitance are interchanged as in Fig. 19.12b, the sum of the currents is

$$C_1\frac{dv_1}{dt} + \frac{v_o}{R} = 0 \tag{19-24}$$

Solving,

$$v_o = -RC_1\frac{dv_1}{dt} \tag{19-25}$$

and the output voltage is proportional to the derivative of the input. For practical reasons involving instability and susceptibility to noise, the differentiator is not so useful as the integrator. (See Example 4.)

Low-Pass Filter

A *passive filter* is a frequency-selective network consisting of passive resistors, inductors, and capacitors. (See Figs. 7.9, 7.11, and 12.15.) Amplifiers can be used to realize *active filters* that avoid the use of bulky and expensive inductors and have other advantages. Low-cost op amps with desirable input and output properties are widely used in active filters.

Example 4

Predict the output voltage of the circuit shown in Fig. 19.13 where the block represents an ideal amplifier with A very large.

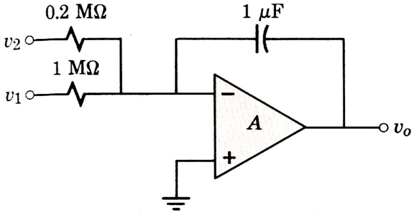


Figure 19.13 Integrator application.

For A very large, v_i is very small and representation as an integrator is accurate. By the principle of superposition,

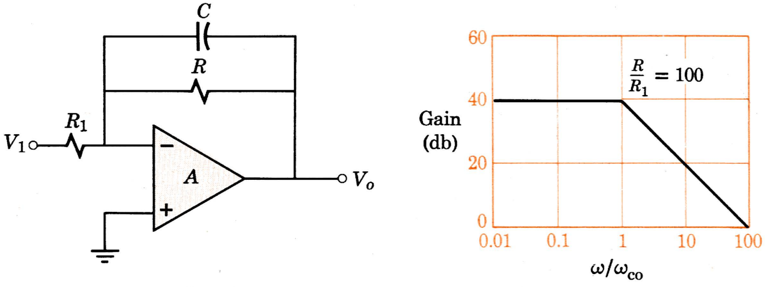
$$v_o = -\frac{1}{C}\int\left(\frac{v_1}{R_1} + \frac{v_2}{R_2}\right) dt$$

$$= -\frac{1}{CR_1}\int\left(v_1 + v_2\frac{R_1}{R_2}\right) dt$$

$$= -\int (v_1 + 5v_2)\, dt$$

The output is the integral of a weighted sum.

As a simple example, consider Fig. 19.14. This is a form of the general inverting circuit with impedance elements $\mathbf{Z}_F$ and $\mathbf{Z}_1$ in the feedback network. Here $\mathbf{V}_1$ is a sinusoidal signal of variable frequency or a combination of signals of various frequencies. The voltage gain (a function of frequency ω) is



(a) Circuit (b) Frequency response

Figure 19.14 Simple low-pass filter.

$$A_F = \frac{\mathbf{V}_o}{\mathbf{V}_1} = -\frac{\mathbf{Z}_F}{\mathbf{Z}_1} = -\frac{\dfrac{1}{(1/R) + j\omega C}}{R_1} = -\frac{R/R_1}{1 + j\omega RC} \tag{19-26}$$

The cutoff or half-power frequency is defined by $\omega_{co}RC = 1$ or

$$\omega_{co} = \frac{1}{RC} \tag{19-27}$$

The voltage gain decreases rapidly above ω_{co} and this serves as a *low-pass filter*.

The response curve of Fig. 19.14b does not exhibit a sharp cutoff. More complicated circuits designed by more sophisticated methods and using several op amps can provide sharp high-, low-, or band-pass filtering. In designing active filters for high-frequency applications (above 1 MHz), the frequency response of the op amp itself must be taken into account.

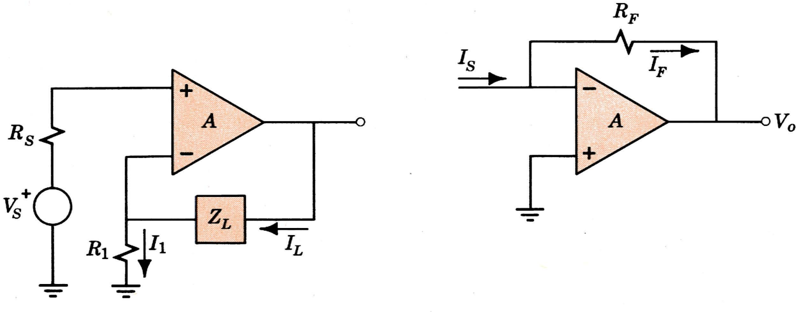
Current-Voltage Converters

The combination of very high gain and feedback results in a pair of useful op amp properties:

1. The voltage across the input terminals, v_i, is practically zero.
2. The current into either input terminal, v_i/R_i, is practically zero.

These properties contribute to the versatility of the op amp in a variety of applications.

One common application is in driving a low-resistance load (a coil, for example) with a voltage source of high internal resistance. The *voltage-to-current converter* shown in Fig. 19.15a performs this function effectively. Since no current flows into the op-amp terminal, there is no drop across R_s and the full voltage V_s appears between the + terminal and ground. Therefore, current $I_1 = V_s/R_1$. Since no current flows into the op amp, current I_1 must be supplied by the feedback current I_L. Therefore,



(a) Voltage to current (b) Current to voltage

Figure 19.15 Voltage-current converter circuits.

$$I_L = I_1 = \frac{V_s}{R_1} \tag{19-28}$$

and the source voltage has been converted into a load current.

The inverse function is performed by the circuit of Fig. 19.15b. Here a given current I_s is introduced into the − terminal of the op amp, setting up an equal current in the feedback resistor R_F. Since the amplifier input voltage v_i is practically zero, the output voltage is

$$V_o = -I_F R_F = -I_s R_F \tag{19-29}$$

and the source current has been converted into a voltage. The current-to-voltage converter is useful where a current is to be measured without introducing an undesired resistance ($R_{oF} \cong 0$) or where a current source, such as a photocell, has a high shunt resistance. (See Example 5.)

Example 5

A photodiode generates 0.2 μA of current per μW of radiant power for a constant reverse-bias voltage. Design a circuit to measure radiant power with a voltmeter.

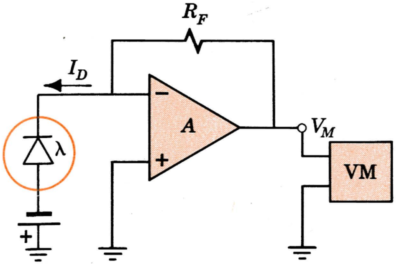


Figure 19.16 Photodiode lightmeter.

The photocurrent could be allowed to flow through a resistance to develop a measurable voltage, but this voltage would reduce the reverse bias and alter the photodiode characteristic.

Instead, we use the circuit of Fig. 19.16. Since $v_i \cong 0$, there is no bias voltage change with a change in I_D.

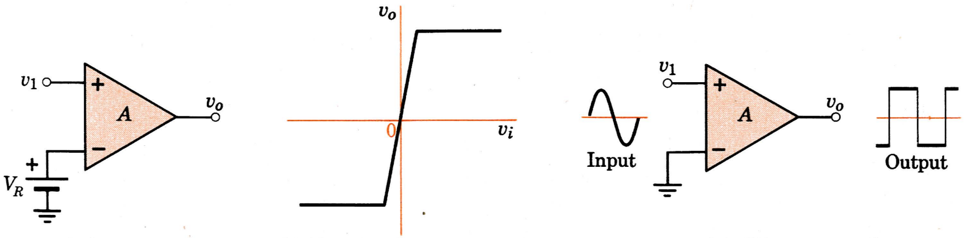
If $R_F = 5\ \text{k}\Omega$ and the radiant power is 100 μW, the voltmeter will read

$$V_M = I_D R_F = (100 \times 0.2) \times 10^{-6} \times 5 \times 10^3$$
$$= 100\ \text{mV}$$

and the scale factor of the "lightmeter" is 1 μW/mV.

Waveform Generators

The ideal operational amplifier is a linear device in that the output is directly proportional to the input for all values of input voltage. There are several important nonlinear applications of op amps, one of the simplest is the *comparator* diagramed in Fig. 19.17a. If the input voltage v_1 is larger than the reference



(a) Comparator (b) Transfer characteristic (c) Clipper operation

Figure 19.17 Nonlinear applications of an operational amplifier.

V_R, the output voltage v_o is positive. Since the gain is very large, only a small voltage difference will drive the amplifier into *saturation*, since the maximum output swing is limited by the supply voltage. (See Fig. 14.10.) The transfer characteristic indicates that a small decrease in v_i (a fraction of a millivolt) will drive the op amp from positive saturation to negative saturation (tens of volts).

If $V_R = 0$, this becomes a *zero-crossing comparator* (Fig. 19.17b). Such a comparator can be used to convert a sine wave into a square wave by the clipping action shown in Fig. 19.17c.

A simple and inexpensive *square-wave generator* can be constructed with one op amp and a pair of back-to-back Zener diodes. In Fig. 19.18a, a capacitor C is charged through resistor R_F from output voltage v_o limited to $+V_Z$ or $-V_Z$ by the diodes in combination with R_S. The op amp compares v_1 to $\frac{1}{2}V_Z$ obtained from the voltage divider where $R_2 = R_3$ (in this case). When $v_i = \frac{1}{2}V_Z - v_1$ changes sign, v_o changes sign. Half of v_o is fed back (positive feedback) to the noninverting terminal to drive the op amp into saturation.

Example 6

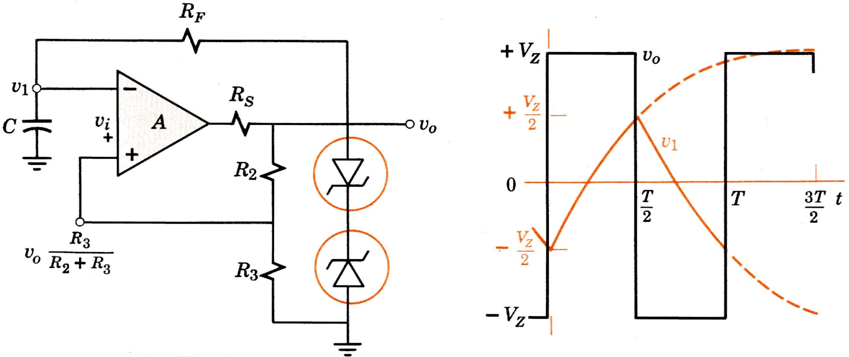
Design a simple square-wave generator for operation at 200 Hz ($T = 0.005$ s) with an amplitude of ±10 V.

Two 10-V diodes (1N4104) are connected back-to-back as in Fig. 19.18a. Assuming op amp operation at ±15 V, approximately $15 - 10 = 5$ V is to be dropped across R_S. Therefore, for $I_Z = 1$ mA, say,

$$R_S = V/I_Z = 5/0.001 = 5000\ \Omega$$

Selecting $C = 0.1\ \mu\text{F}$, $R_2 = R_3 = 1\ \text{M}\Omega$, $H = 0.5$.

$$R_F = \frac{T}{2C\ln 3} = \frac{5 \times 10^{-3}}{2 \times 10^{-7} \times 1.1} = 22{,}750 \cong 22\ \text{k}\Omega$$



(a) Circuit

(b) Waveforms for $R_2 = R_3$

Figure 19.18 An elementary square-wave generator.

To see how the circuit works, assume that $R_2 = R_3$ and that $V_Z = 10$ V. At $t = 0^-$, v_1 is approaching -5 V. At $t = 0$, v_1 reaches -5.01 V, say, and v_i goes positive, sending the op amp toward positive saturation but limited to $+10$ V. Since v_1 is the voltage on C, it cannot suddenly change and at $t = 0^+$, $v_1 \cong -5$ V. Since $v_o = +10$ V, the voltage now tending to force current through R_F is $v_o - v_1 = 10 - (-5) = 15$ V. The capacitor voltage will rise exponentially, or

$$v_1 = 15(1 - e^{-t/R_FC}) - 5 \tag{19-30}$$

When v_1 exceeds $+\frac{1}{2}V_Z = +5$ V, the + input terminal is less positive than the − terminal, v_i changes sign, and v_o goes negative. Half of v_o is fed back to make v_i more negative, and v_o is driven rapidly and solidly to $-V_Z$.

In general, where $R_3/(R_2 + R_3) = H$ and $v_1 = -HV_Z$ at $t = 0$,

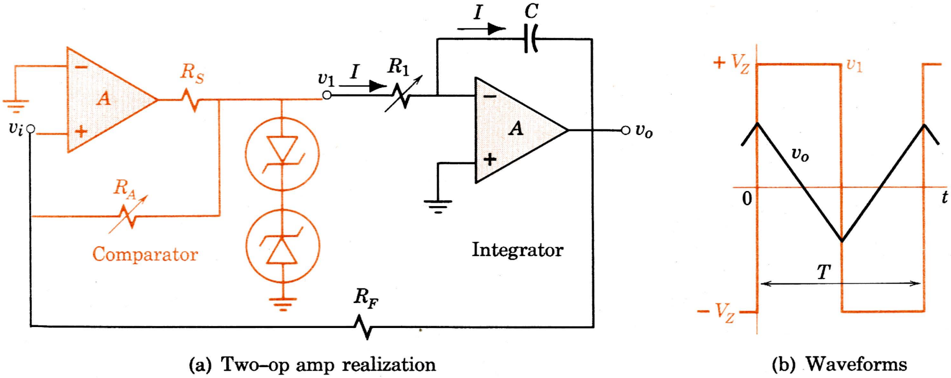
$$v_1 = (1 + H)V_Z(1 - e^{-t/R_FC}) - HV_Z \tag{19-31}$$

for the first half cycle. At $t = T/2$, $v_1 = +HV_Z$. Substituting these values in Eq. 19.31 and solving, the period is

$$T = 2R_FC \ln \frac{1 + H}{1 - H} \tag{19-32}$$

Such a square-wave generator with symmetric Zener diodes and stable network elements will provide good square waves in the audiofrequency range.

Circuits containing a few operational amplifiers can be used to generate pulse trains, ramps, sawtooths, sine waves, or almost any other waveform of interest. For example, consider the *triangular-wave generator* in Fig. 19.19. In response to changes in v_i, the comparator switches between positive and negative saturation with its output v_1 clamped at $+V_Z$ or $-V_Z$. Assume that $v_1 = +V_Z$ at $t = 0$. The current flowing into the integrator is $I = V_Z/R_1$, which also serves



(a) Two-op amp realization (b) Waveforms

Figure 19.19 An elementary triangular-wave generator.

to charge capacitor C. The integrator output v_o is just the capacitor voltage or

$$v_0 = V_0 - \frac{1}{C}\int_0^t I\,dt = V_0 - \frac{I}{C}t \tag{19-33}$$

This is a negative-slope *ramp* voltage. A part of v_o is fed back through R_F to the + terminal of the comparator. When v_i changes sign to negative, the comparator switches to negative saturation, v_1 switches to $-V_Z$ and constant current I is reversed. This initiates a positive-going ramp and v_o is triangular. The amplitude can be controlled by R_A, which adjusts the feedback factor, and the frequency can be adjusted by R_1, which controls the capacitor charging current.

Practical Considerations

The ideal operational amplifier is characterized by infinite gain, infinite bandwidth, infinite input impedance, and zero output impedance. By the ingenious use of technological advances, design engineers provide at incredibly low cost commercial amplifiers that come close to the ideal.

A Four-Stage Op Amp

To provide the desired characteristics, several stages are required. In Fig. 19.20, the first stage is a *differential amplifier* with a double-ended output; this provides high input resistance, high gain for *difference* signals, and rejection of signals *common* to both terminals. The second stage is a single-output differential amplifier that provides more gain and more discrimination. The third stage is an *emitter follower* to provide low output resistance and isolation of the amplifier from the load. The last stage is a combination *level shifter* and *driver* stage. The level shifting corrects for any dc offsets that have been introduced by bias networks or by component imbalances. The driver is a power amplifier to provide large output currents from a low-output resistance.

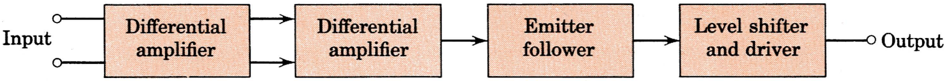
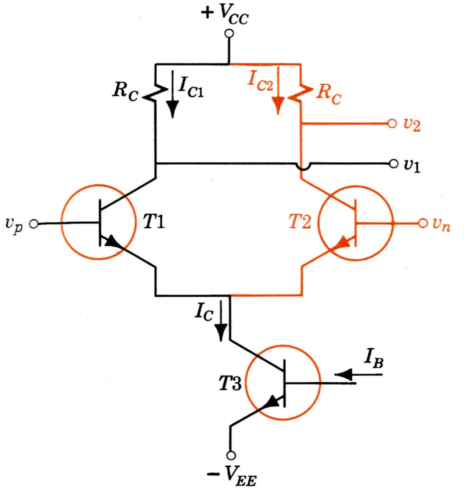


Figure 19.20 Block diagram of a four-stage operational amplifier.

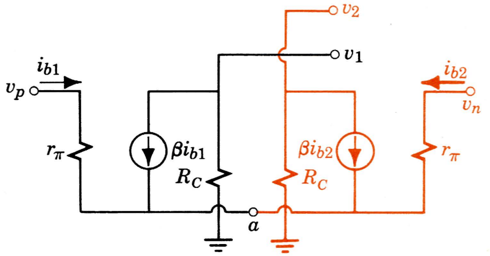
The differential amplifier† is a versatile component of many electronic instruments and deserves further consideration. As shown in Fig. 19.21a, transistor T3 is a constant-current source where $I_{C3} = \beta I_{B3}$ and I_{B3} is obtained from a biasing network. The remainder of the circuit is a symmetric amplifier of matched resistors and transistors; in integrated circuit form, the corresponding elements

---

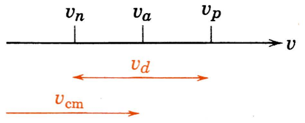
† Here we refer to an electronic circuit, a portion of an IC op amp; the entire op amp can also function in a differential amplifier circuit (Fig. 19.9).



(a) Circuit



(b) Small-signal model



(c) Signal representation

Figure 19.21 Analysis of a differential amplifier.

are fabricated simultaneously and are nearly identical in all respects. Qualitatively speaking, any increase in v_p increases I_{B1}, which increases I_{C1} and lowers v_1. Since $I_{C1} + I_{C2} = I_{C3} =$ constant, an increase in I_{C1} results in a corresponding decrease in I_{C2} and v_2 goes up. The output $v_o = v_2 - v_1$ responds to the *difference* in v_p and v_n. If v_p and v_n both increase due to a *common* signal, T1 and T2 would have to respond in identical fashion, but I_{C1} and I_{C2} are unaffected since their sum must be constant. Therefore, $v_2 = v_1$ and there is no output.

For a quantitative analysis, we replace the transistors by their small-signal models as in Fig. 19.21b. Summing the currents into node a, we write

$$\Sigma i = (1+\beta)i_{b1} + (1+\beta)i_{b2} = (1+\beta)\frac{v_p - v_a}{r_\pi} + (1+\beta)\frac{v_n - v_a}{r_\pi} = 0 \tag{19-34}$$

Hence, $(v_p - v_a) + (v_n - v_a) = 0$ which yields

$$v_a = \frac{v_p + v_n}{2} \tag{19-35}$$

or v_a takes on a value just halfway between v_p and v_n (Fig. 19.21c). Now let the input signals be described in terms of a common-mode signal v_{cm} and a difference signal v_d so that $v_p = v_{\mathrm{cm}} + v_d/2$ and $v_n = v_{\mathrm{cm}} - v_d/2$. Using the small-signal models,

$$v_1 = -\beta\frac{v_p - v_a}{r_\pi}R_C = -\frac{\beta R_C}{r_\pi}(v_{\mathrm{cm}} - v_a + v_d/2)$$

$$v_2 = -\beta\frac{v_n - v_a}{r_\pi}R_C = -\frac{\beta R_C}{r_\pi}(v_{\mathrm{cm}} - v_a - v_d/2)$$

and the output voltage is

$$v_o = v_2 - v_1 = \frac{\beta R_C}{r_\pi} v_d = \frac{\beta R_C}{r_\pi}(v_p - v_n) \qquad (19\text{-}36)$$

The common-mode signal is rejected and the difference signal is amplified.

Common-Mode Rejection Ratio

Ideally, a differential amplifier responds only to the difference voltage $v_p - v_n$ and not to the common-mode voltage v_{cm}. Practically, however, an op amp with an input v_{cm} produces some output v_{ocm}; the ratio

$$A_{\mathrm{cm}} = \frac{v_{\mathrm{ocm}}}{v_{\mathrm{cm}}} \qquad (19\text{-}37)$$

is called the *common-mode gain.* The ratio of differential gain A to A_{cm} is called the common-mode rejection ratio, CMRR, where

$$\mathrm{CMRR} = \frac{A}{A_{\mathrm{cm}}} \qquad (19\text{-}38)$$

or, in decibels, CMRR(dB) = 20 log CMRR. Typical values run from 50 to 100 dB.

Example 7

A 741 op amp has a gain $A = 2 \times 10^5$ and a CMRR of 90 dB. The input consists of a difference signal $v_p - v_n = 10\,\mu\mathrm{V}$ and a common-mode signal 100 times as large, i.e., $v_p = 1005\,\mu\mathrm{V}$ and $v_n = 995\,\mu\mathrm{V}$. Determine A_{cm} and the output voltage.

By Eq. 19-38, the common-mode gain is

$$\begin{aligned} A_{\mathrm{cm}} &= A/10^{90/20} = 2 \times 10^5 \times 10^{-4.5} \\ &= 2 \times 10^{0.5} = 6.3 \end{aligned}$$

By superposition, the output voltage is

$$\begin{aligned} v_o &= A_d v_d + A_{\mathrm{cm}} v_{\mathrm{cm}} \\ &= 2 \times 10^5 \times 10^{-5} + 6.3 \times 10^{-3} \\ &= 2.0063 \text{ V} \end{aligned}$$

Even in this extreme case, imperfect CMR introduces only a 0.3% error in the output.

Input Offset Voltage

When both input terminals are grounded, the ideal operational amplifier develops zero output voltage. Under the same conditions, a practical amplifier will have a finite output because of inevitable small imbalances in the components. By definition, the *input offset voltage* V_{OS} is the dc input voltage required to reduce the output to zero. A practical op amp can be modeled by an ideal op amp with an input voltage source equal to V_{OS}, typically a few millivolts (Fig. 19.22). For the configuration shown with $v_1 = 0$, V_{OS} is equivalent to a signal applied to the noninverting terminal. Therefore, the output voltage will be

$$V_o = \frac{R_1 + R_F}{R_1} V_{OS} \qquad (19\text{-}39)$$

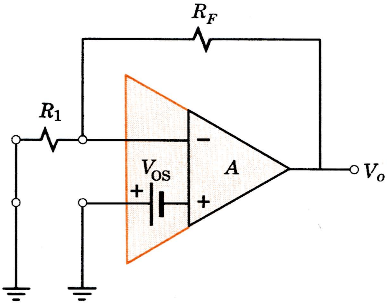


Figure 19.22 An op amp with input voltage offset.

The offset voltage, and its "drift" with changes in temperature and time, is one of the important sources of error in operational amplifier circuits. Op amps with appropriately low values of offset voltage must be selected or special nulling circuits must be employed. Many IC op amps provide connections to internal points for this purpose.

Bias Current Offset

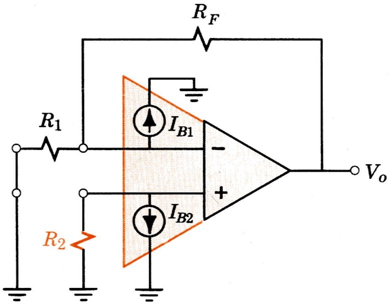


Figure 19.23 An op amp with bias current offset.

The input terminals of a differential amplifier (Fig. 19.21a) carry the base bias currents for the transistors in the input stage. In an op amp circuit, these dc currents (typically a fraction of a μA) produce IR voltages that appear in the output as spurious signals. A practical op amp can be modeled by an ideal op amp with current generators in each lead (Fig. 19.23). In an IC op amp, bias currents I_{B1} and I_{B2} will be very nearly equal and their effects can be balanced out by the insertion of resistor R_2.

By superposition, the total effect can be predicted as the sum of the effects of each current generator considered separately. Since V_{R1} is essentially zero, all I_{B1} flows through R_F and its contribution to output voltage is

$$V_{o1} = I_{B1} R_F \qquad (19\text{-}40)$$

The Thévenin equivalent of I_{B2} and R_2 is a signal $V_T = -I_{B2}R_2$ at the noninverting terminal; its contribution is

$$V_{o2} = \frac{R_1 + R_F}{R_1}(-I_{B2}R_2) \qquad (19\text{-}41)$$

Hence the output offset due to bias currents is

$$V_o = V_{o1} + V_{o2} = I_{B1}R_F - I_{B2}R_2(1 + R_F/R_1) \qquad (19\text{-}42)$$

To minimize the effects of the nearly equal bias currents, we equate their coefficients and solve for R_2, obtaining

$$R_2 = \frac{R_1 R_F}{R_1 + R_F} \tag{19-43}$$

For this value of R_2, the remaining output offset is

$$V_o = (I_{B1} - I_{B2})R_F = I_{OS}R_F \tag{19-44}$$

where I_{OS} is the *input offset current,* which is usually much smaller than average bias current $(I_{B1} + I_{B2})/2$. As indicated in Example 8, input offset current is important in amplifier circuits with large resistance values. Input offset voltage is important in circuits with high gain.

Example 8

An op amp with the following "typical" specifications:

Input bias current = 0.1 μA
Input offset current = 0.02 μA
Input offset voltage = 1 mV

is used in an inverting amplifier with $R_1 = 20\ \text{k}\Omega$ and $R_F = 1\ \text{M}\Omega$.

Specify the proper value of balancing resistor R_2. Then predict the output offset due to bias current without R_2 and with R_2. Predict the output offset due to input offset voltage.

By Eq. 19-43,

$$R_2 = \frac{R_1 R_F}{R_1 + R_F} = \frac{(2 \times 100) \times 10^8}{(2 + 100) \times 10^4} = 19.6\ \text{k}\Omega$$

Without R_2, bias current I_{B1} produces

$$V_o = I_{B1}R_F = 1 \times 10^{-7} \times 1 \times 10^6 = 0.1\ \text{V}$$

With R_2, the input offset current produces

$$V_o = I_{OS}R_F = 2 \times 10^{-8} \times 1 \times 10^6 = 0.02\ \text{V}$$

By Eq. 19-39, the input offset voltage produces

$$V_o = V_{OS}(R_1 + R_F)/R_1$$
$$= 10^{-3}(1.02 \times 10^6)/2 \times 10^4 = 0.051\ \text{V}$$

Frequency Response

Whereas the ideal operational amplifier has an infinite bandwidth, the gain A of a practical op amp begins to drop off as the frequency increases. In many applications, however, the frequency response may be considered *flat,* because the use of feedback greatly increases the bandwidth of the circuit. As indicated by Example 8 in Chapter 18, the bandwidth is increased by the same factor $1 - AH$ as the gain is reduced. For an op amp with $A = 10^5$ used in a circuit where $A_F = 100$, the bandwidth of the circuit (as defined by the 3 db frequency) will be $A/A_F = 1000$ times as great as the bandwidth of the op amp alone. Alternatively, the *gain-bandwidth product* (Eq. 18-44) is equal to the *unity-gain frequency,* or

$$A_F f_{2F} = f_T \tag{19-45}$$

If f_T is supplied by the manufacturer, the bandwidth f_{2F} is determined for any gain A_F.

Stability

Actually, the high-frequency gain of an op amp is usually limited on purpose by the designer. At high frequencies, inherent capacitive effects introduce phase shift of the output signal with respect to the input. If this phase shift equals 180°, the feedback will be *positive* and oscillation is possible if the loop gain AH is equal to 1 (see p. 484).

One way to ensure that the amplifier is free from oscillation, or *stable*, is to include an RC frequency compensation network. Such a network is essentially a low-pass filter (Fig. 19.14) and introduces a maximum of 90° phase shift and a *gain rolloff* of 20 dB per decade or 6 dB per octave. This gain characteristic is superimposed on the characteristic of the basic amplifier and reduces the gain to unity before a 180° phase shift can occur. In an op amp, frequency compensation may be internal or connections may be brought out for external compensation.

Slew Rate

The bandwidth determines the ability of an operational amplifier to follow rapidly changing small signals. In addition, there are limitations on the rate at which the output can follow large signals. This results from the fact that only finite currents are available within the amplifier to charge the various internal capacitances. The maximum rate at which the output voltage can swing from most positive to most negative is the *slew rate* ρ, typically a few volts per μs. For example, for an op amp with a slew rate of 1 V/μs operating at $V_{CC} = \pm 15$ V, the switching time will be

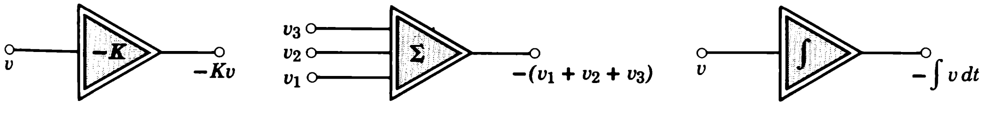
$$T_{\text{sw}} \cong \frac{\text{supply voltage}}{\text{slew rate}} = \frac{30}{1} = 30\,\mu\text{s}$$

Slew rate may be a limitation on frequency response to large sinusoids, but it is of primary importance in switching applications.

The Analog Computer

The modern electronic analog computer is a precision instrument. Its basic purpose is to predict the behavior of a physical system that can be described by a set of algebraic or differential equations. The programing procedure is to arrange the operational amplifiers to perform the operations indicated in the describing equations and provide a means for displaying the solution.

In addition to the operational amplifiers (identified by function in Fig. 19.24),



(a) Inverting amplifier (b) Adder (c) Integrator

Figure 19.24 Symbolic representation of operational amplifier functions.

the practical computer includes an assortment of precision resistors and capacitors, a function generator to provide various inputs, means for introducing initial conditions, potentiometers for introducing adjustable constants, switches for controlling the operations, an oscilloscope or recorder for displaying the output, and a problem board for connecting the components in accordance with the program. In the hands of a skillful operator, the analog computer faithfully simulates the physical system, provides insight into the character of the system behavior, and permits the design engineer to evaluate the effect of changes in system parameters before an actual system is constructed.

Solution by Successive Integration

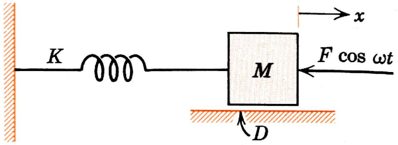


Figure 19.25 A physical system.

In one common application the analog computer is used to solve linear integro-differential equations. To illustrate the approach, let us predict the behavior of the familiar physical system shown in Fig. 19.25. Assuming that mass is constant, that the spring is linear, and that friction force is directly proportional to velocity, the system is described by the linear differential equation

$$\Sigma f = 0 = -F \cos \omega t - M\frac{d^2x}{dt^2} - D\frac{dx}{dt} - \frac{1}{K}x \tag{19-46}$$

and a set of initial conditions. The behavior of the system can be expressed in terms of displacement $x(t)$ or velocity $u(t)$ where $u = dx/dt$. We wish to display this behavior, so we proceed to program the computer to solve the equation.

The first step is to solve for the highest derivative. Anticipating the inversion present in the operational amplifier used, we write

$$\frac{d^2x}{dt^2} = -\left(\frac{F}{M}\cos \omega t + \frac{D}{M}\frac{dx}{dt} + \frac{1}{KM}x\right) \tag{19-47}$$

To satisfy this equation, the mathematical operations required are: addition, integration, inversion, and multiplication by constants. The required addition and the two integrations are shown in Fig. 19.26; in each operation there is an inversion.

The next step is to arrange the computer elements to satisfy the equation. Knowing the required inputs to the adder, we can pick off the necessary signals (in the form of voltages) and introduce the indicated multiplication constants

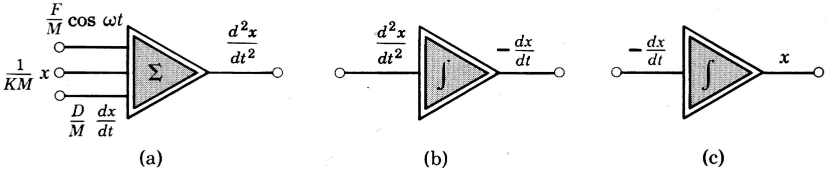


Figure 19.26 Operations needed in solving Eq. 19.47 by successive integration.

and inversions. Ignoring initial conditions for the moment, one possible program is outlined in Fig. 19.27. Closing the circuit imposes the condition that the equation be satisfied. A properly synchronized cathode-ray tube connected at terminal 1 would display the displacement $x(t)$. The velocity $u(t)$ is available at terminal 2, but an inversion would be necessary to change the sign. Here six operational amplifiers are indicated; by shrewd use of amplifier capabilities, the solution can be accomplished with only three operational amplifiers.

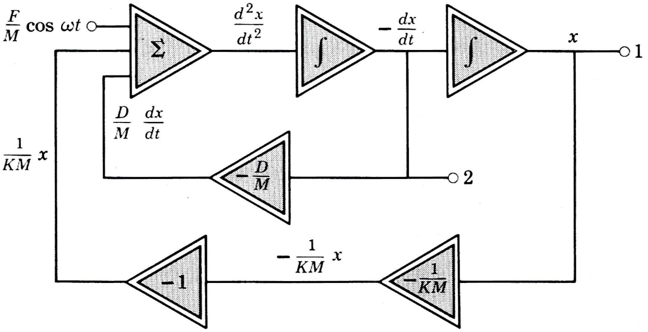


Figure 19.27 Analog computer program for Eq. 19-47.

Example 9

For the physical system in Fig. 19.25 mass $M = 1$ kg, friction coefficient $D = 0.2$ N-s/m, and compliance $K = 2$ m/N. Devise an analog computer program using only three operational amplifiers to obtain the displacement and velocity for an applied force $f = F\cos\omega t$ N. (Note: Assume the op amps are single-input inverting.)

From Example 4 we know that one operational amplifier can add and integrate the weighted sum. In this case, $M = 1$, $D/M = \frac{1}{5}$, and $1/KM = \frac{1}{2}$; after one integration, Eq. 19-47 becomes

$$\frac{dx}{dt} = -\int\left(F\cos\omega t + \frac{1}{5}\frac{dx}{dt} + \frac{1}{2}x\right)dt \qquad (19\text{-}48)$$

A convenient reference is to let $RC = 1$ and specify resistances in megohms and capacitances in microfarads. For $C = 1\ \mu$F and $R_1 = 1$ MΩ, a pure integration is performed; for $R_2 = 0.2$ MΩ, v_2 has a weighting of 5 in the result. On that basis the appropriate values for Eq. 19-48 are as shown in Fig. 19.28.

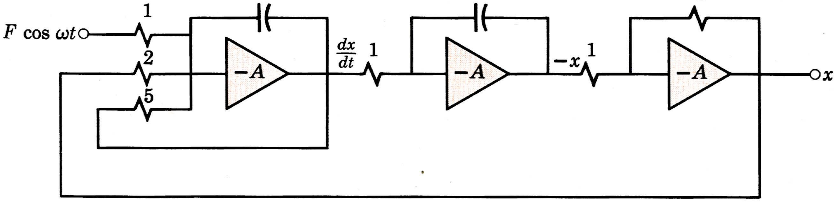


Figure 19.28 Analog computer program for solving Eq. 19-48. Resistance in megohms (MΩ) and capacitance in microfarads (μF).

The Practical Computer

In the practical computer there are provisions and limitations that we have not considered. One provision is for inserting a variety of initial conditions. How could we impose the condition that the initial displacement in Example 9 is $X_0 = 10$ m? In these programs, displacement is represented by a voltage at a particular point in the circuit. A convenient place to put an initial voltage is on a capacitor. Figure 19.13 indicates that, since v_i is negligibly small because of the high gain A, the voltage across the 1-μF capacitor is just equal to v_o. An initial voltage of 10 V on the capacitor of the second integrator of Fig. 19.28 is analogous to an initial displacement of 10 m. The practical computer provides voltage sources and potentiometers (variable voltage dividers, see Fig. 19.29) to introduce initial displacements or velocities by closing the *reset* switch. To *compute* the switch is opened.

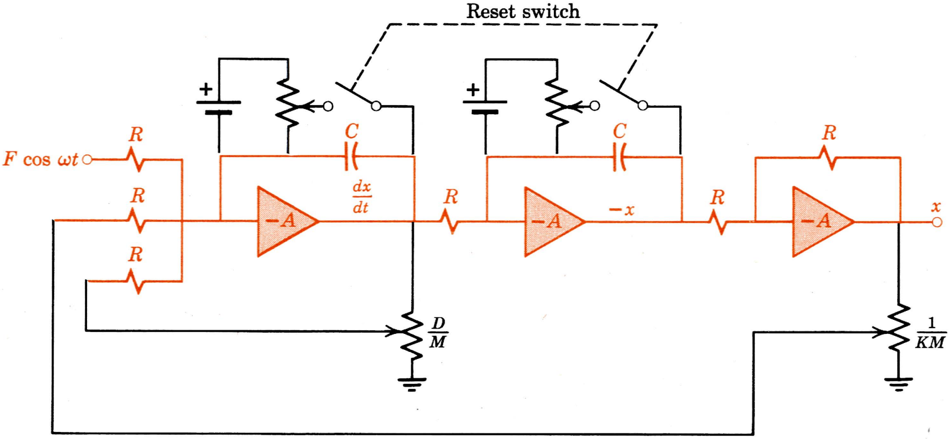


Figure 19.29 A practical computer with provision for introducing initial conditions.

Potentiometers also provide flexibility in the form of adjustable multiplying constants. Using calibrated potentiometers, the designer has a continuous range of parameter values at his fingertips. If the multiplying constants are greater than unity, the value of R in Fig. 19.29 can be changed to allow the use of potentiometers.

Another advantage of the analog computer is that the time scale can be changed at will. In Example 9, 1 second of computer time corresponds to 1 second of real time. To slow down an event that takes only a few milliseconds, we can arrange the *time scale* so that 1 millisecond of real time corresponds to 1 second of computer time. To change the time scale by a factor a, we can let the computer time variable $t_c = at$ and rewrite the differential equation. Increasing the integrator time constant RC by a factor a slows down the computer by the same factor. Changing the time scale changes the voltage levels produced by integration and it may be necessary to modify the *amplitude scale.*

One limitation of the practical computer is that the amplifiers are linear over only a finite voltage range, usually around ±10 V. To ensure accuracy and to avoid possible damage, the amplifier outputs should never exceed the rated maximum value. On the other hand, if the signal voltage is too low at any stage, it may be masked by the *noise* voltages that are present in any electronic equipment. (One reason for preferring integration over differentiation in analog computation is that random noise voltages may have large time derivatives, while their integrated values are usually zero.) Once the time-scale factor has been established, an amplitude-scale factor can be introduced to keep signal voltages as large as possible without exceeding the rated values.

Example 10

For the physical system in Fig. 19.25 mass $M = 1$ kg, friction coefficient $D = 2$ N-s/m, and compliance $K = 0.0001$ m/N. Plan an analog computer program to obtain a record showing displacement and velocity for an applied force defined by $f = -200 \cos 200t$ N. Note: the available pen recorder will not follow variations above about 2 Hz.

With this very stiff spring the governing equation becomes

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 10{,}000x = -200 \cos 200t$$

The frequency of the applied force is 200 rad/s or 32 Hz. Reasoning by analogy from Eq. 7-27, we see that the undamped natural frequency of the system is

$$\omega_n = \frac{1}{\sqrt{KM}} = \frac{1}{\sqrt{0.0001}} = 100 \text{ rad/s or } 16 \text{ Hz}$$

Because the pen recorder available will not follow such rapid variations, we wish to expand the time scale by a factor of 10. Letting computer time $t_c = 10t$, $u = dx/dt = 10\,dx/dt_c$ and $d^2x/dt^2 = 100\,d^2x/dt_c^2$. In terms of computer time,

$$100\frac{d^2x}{dt_c^2} + 20\frac{dx}{dt_c} + 10{,}000x = -200 \cos 20t_c$$

or

$$\frac{d^2x}{dt_c^2} + 0.2\frac{dx}{dt_c} + 100x = -2 \cos 20t_c$$

The program for such a computation proceeds as in Example 9 (see Fig. 19.28). The resulting plot of displacement $x(t_c)$ must be interpreted in terms of the changed time variable.

The plot of velocity in Example 10 must be interpreted also, since u_t is equal to $10u_c$. One possibility is to adjust the amplitude scale so that the output of the first integrator is $10\,dx/dt_c$ instead of dx/dt as shown in Fig. 19.28. Since the force, velocity, and displacement are all represented by voltages, provision must be made to place these at proper levels; in complicated systems, amplitude scaling becomes difficult. Fortunately, most commercial computers have overload indicators on each amplifier; the variable to any amplifier that shows overloading can be rescaled.

An interesting application of the analog computer is as a *function generator.* A desired time function can be obtained as a solution to another equation. For example (see Problem 6), with the proper initial conditions the function $y = Y \cos \omega t$ can be obtained as the solution of the equation

$$\frac{d^2y}{dt^2} + \omega^2 y = 0$$

SUMMARY

- The availability of low-cost high-performance op amps has radically altered the design of electronic signal processing equipment.
- The ideal op amp has infinite gain, bandwidth, input impedance, and common-mode rejection ratio and zero output impedance.
  Negative feedback is used to obtain the desired characteristics.
- Applications include: inverting and noninverting amplifiers, summing circuits, voltage followers, integrators, differentiators, buffers, filters, current-voltage converters, and waveform generators.
  The unity-gain buffer is used to isolate a source from a load.
  Active filters with improved characteristics can be realized with op amps.
  Voltage signals can be converted to current signals and vice versa.
  Various waveforms can be generated by op amps operating nonlinearly.
- The differential amplifier is a symmetric two-input amplifier that provides high differential gain and a high common-mode rejection ratio.
  Practical limitations on op amp performance include: input offset voltage, bias current offset, frequency response, stability, and slew rate.
- The analog computer is a flexible model of the system being studied; system behavior is represented by continuously varying quantities.
  Programing consists in arranging the operational amplifiers to perform efficiently the operations indicated in the equations describing the system.
  Integration is preferable to differentiation for practical reasons.
- In a practical computer there are provisions for introducing initial conditions and for adjusting parameter values.
  Changes in time and amplitude scale may be necessary.

REFERENCES

1. Roger Melen and Harry Garland, *Understanding IC Operational Amplifiers*, Howard W. Sams & Co., Indianapolis, 1971.

   An excellent nonmathematical introduction to IC op amps; how they are made, how they work, and how they are used. Practical applications.
2. Jerry Eimbinder, ed., *Application Considerations for Linear Integrated Circuits*, Wiley-Interscience, New York, 1970.

   Practical advice for design engineers selecting and applying IC op amps. Specifications, general applications, and design considerations.
3. William Oldham and Steven Schwarz, *An Introduction to Electronics*, Holt, Rinehart, and Winston, New York, 1972.

   Includes a detailed analysis of differential and operational amplifiers and their applications at the undergraduate level.
4. Jerald Graeme and Gene Tobey, editors, *Operational Amplifiers: Design and Application*, McGraw-Hill Book Co., New York, 1971.

   Thorough treatment of design and applications of IC op amps; prior circuit theory required. Appendix includes concise summary of basic concepts.

REVIEW QUESTIONS

1. What is an operational amplifier?
2. List the specifications of an ideal op amp.
3. Why is the gain of an amplifier circuit independent of the gain of the op amp employed?
4. In circuit analysis of an ideal op amp with feedback, what values are assumed for v_i and i_i? Why?
5. How can an op amp be used to add two analog signals?
6. What is a voltage follower? How does it work?
7. Why is a differential amplifier used in the input stage of an op amp?
8. Distinguish between difference and common-mode signals.
9. How does an op amp integrator work? A differentiator?
10. In Fig. 19.15a, little current is drawn from the source; where does I_L come from?
11. Explain the operation of the square-wave generator in Fig. 19.18.
12. Why does input bias current cause output voltage error?
13. Why is high-frequency gain intentionally reduced in an op amp?
14. What is the basic difference between digital and analog computers?
15. Why is the differentiator less useful than the integrator in computation?
16. What is the difference between an analog computer and an analogous circuit on which measurements are taken?
17. Outline the procedure followed in programing an analog computer.
18. What would be needed to permit the solution of nonlinear equations?
19. What are the objectives of time scaling and amplitude scaling?

EXERCISES

1. Determine the error in Eq. 19-3 if $R_1 = R = 1\ \text{M}\Omega$, $R_i = 10\ \text{M}\Omega$, $A = 5 \times 10^4$, and R_o (the output impedance of the amplifier) is negligible.
2. An op amp has the following low-frequency parameters: $R_i = 1\ \text{M}\Omega$, $R_o = 100\ \Omega$, and $A = 10^5$. Design an amplifier circuit with:
   (a) $A_F = -200$, $R_{iF} = 5\ \text{k}\Omega$.
   (b) $A_F = -30$, $R_{iF} = 20\ \text{k}\Omega$.
   (c) $A_F = +20$, $R_{iF} \geq 1\ \text{M}\Omega$.
   (d) $A_F = +50$, $R_{iF} \geq 1\ \text{M}\Omega$.

3. Given the operational amplifier in Fig. 19.30:
   (a) Define v_o in terms of the given inputs.
   (b) Describe the operation performed assuming sinusoidal inputs.

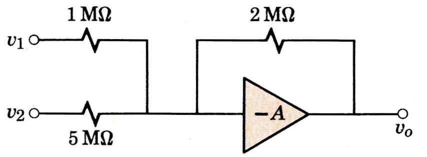


Figure 19.30

4. Using an ideal op amp, design a circuit to produce an output described by:
   (a) $v_o(t) = -5v_1(t) - 2v_2(t)$.
   (b) $v_o(t) = -v_1(t) - 10v_2(t)$.
5. A voltmeter requires 1 mA to provide a full-scale deflection of 10 V. It is to be converted into an electronic voltmeter with an input resistance of at least 10 MΩ and ranges of 0 to 1 and 0 to 10 V.
   (a) Calculate the "input" resistance of the basic instrument.
   (b) Using an ideal op amp, design a circuit to provide the desired characteristics.
   (c) Predict the voltage indicated on the two voltmeters when connected, in turn, across a device represented by $V_T = 6$ V in series with $R_T = 500\ \Omega$.
6. Derive Eq. 19-7 by simultaneous solution of Eqs. 19-6.
7. A mistake is made in connecting up a voltage follower; v_1 is connected to the − terminal and v_o is fed back to the + terminal. Derive an expression for gain A_F and estimate R_i; comment on the usefulness of this circuit.
8. A cheap op amp with $A = 10^4$, $R_i = 10\ \text{k}\Omega$, and $R_o = 1\ \text{k}\Omega$ is used in a unity-gain buffer. Predict R_{iF} and R_{oF}.
9. In a differential amplifier (Fig. 19.9), the input signals are $v_1 = 1.01$ V and $v_2 = 0.99$ V. Predict the output signal v_o if:
   (a) $R_1 = R_2 = 10\ \text{k}\Omega$ and $R_F = R_3 = 100\ \text{k}\Omega$.
   (b) $R_1 = 10\ \text{k}\Omega$, $R_2 = 9\ \text{k}\Omega$, and $R_F = R_3 = 100\ \text{k}\Omega$.
10. The current in a 100-Ω load R_L is to be proportional to $v_1 + 10v_2$, where v_1 and v_2 are sources with effective output resistances of over 10 kΩ. Using op amps of the type described in Exercise 2, design an effective circuit.
11. Repeat Exercise 3 for Fig. 19.31.

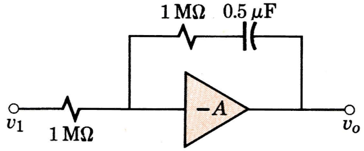


Figure 19.31

12. Derive an equation for v_o/v_1 for Fig. 19.32. What function is performed by the circuit?

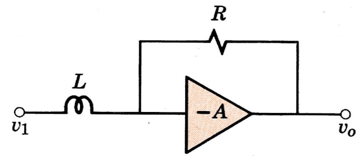


Figure 19.32

13. Devise a sweep circuit using an operational amplifier to provide an output voltage proportional to time. Show a reset switch to ensure that $v_o = 0$ at $t = 0$.
14. The output of a flowmeter is $v = Kq$, where q is in cm³/s and $K = 20$ mV-s/cm³. The effective output resistance of the flowmeter is 2000 Ω. Design a circuit that will develop an output voltage $V_o = 10$ V (to trip a relay) after 200 cm³ have passed the metering point.
15. Design a low-pass filter with a dc gain of 46 dB and a cutoff frequency of 1 kHz.
16. Design a low-pass filter with a dc gain of 60 dB and a cutoff frequency of 100 Hz.
17. Derive an expression for gain of the circuit of Fig. 19.33 as a function of frequency ω. For $R_1 = 10\ \text{k}\Omega$, $C_1 = 1\ \mu\text{F}$, $C = 0.1\ \mu\text{F}$, and $R = 100\ \text{k}\Omega$, determine the critical frequencies and sketch the gain as in Fig. 19.14b. What function is performed at low frequencies? At high frequencies?

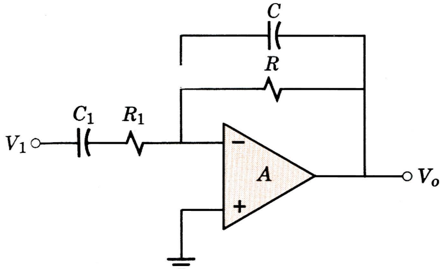


Figure 19.33

18. The op amp of Exercise 2 is used in the circuit of Fig. 19.15b. Derive a general expression for input resistance R_{iF} and evaluate it for the given op amp parameters and $R_F = 1\ \text{k}\Omega$.
19. (a) Analyze the circuit of Fig. 19.34 and derive an expression for i_L in terms of the given quantities.

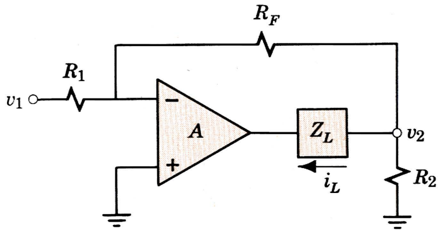


Figure 19.34

(b) A source characterized by $v_s = 2$ V in series with $R_s = 1\ \text{k}\Omega$ is connected to the v_1 terminal. If $R_1 = 100\ \text{k}\Omega$, $R_F = 1\ \text{M}\Omega$, and $R_2 = 10\ \text{k}\Omega$, determine i_L. What function is performed by this circuit and what are its virtues?

20. (a) Derive Eq. 19-32 for the square-wave generator.
    (b) For convenience, it is desired to have the period $T = 2R_FC$; specify the value of H.
21. In Fig. 19.19a, $V_Z = 10$ V, $R_S = 5\ \text{k}\Omega$, $R_A = R_F = 100\ \text{k}\Omega$, $R_1 = 100\ \text{k}\Omega$, and $C = 0.1\ \mu\text{F}$. Assume that at $t = 0$, $v_1 = V_Z$, and $v_o = 0$ V; at what time t will v_i change sign? Determine the amplitude and the frequency of the triangular wave.
22. Determine the CMRR for the differential amplifier of Exercise 9b.
23. An IC differential amplifier (Fig. 19.21) operates with $I_C = 200\ \mu\text{A}$. Assuming $\beta = 100$, estimate the input resistance $R_i = (v_p - v_n)/i_{b1}$ at room temperature.
24. For the differential amplifier of Fig. 19.21, $\beta = 99$, $r_\pi = 20\ \text{k}\Omega$, $R_{C1} = 100\ \text{k}\Omega$ and $R_{C2} = 110\ \text{k}\Omega$. Estimate the differential gain, the common-mode gain, and the CMRR.
25. The op amp of Exercise 2 has an input bias current of $1\ \mu\text{A}$ and an input offset current of $0.1\ \mu\text{A}$.
    (a) Design an amplifier with a gain of +20, assuming $R_F = 100\ \text{k}\Omega$.
    (b) Design an amplifier with a gain of −20, assuming $R_F = 100\ \text{k}\Omega$.
26. A 741 op amp has a low-frequency gain of 10^5 and a unity-gain frequency of 1 MHz.
    (a) Estimate the upper cutoff frequency and plot gain in dB versus frequency on a log scale.
    (b) Plot on the same graph, the curve for an inverting amplifier with a low frequency gain of −100 using the 741.
27. Draw block diagrams of computer programs to solve:

(a) $\begin{cases} x - 3y = 5 \\ x + 2y = \sin \omega t \end{cases}$

(b) $\begin{cases} x + 2y = t \\ x - 3y = 0 \end{cases}$

28. Draw block diagrams of efficient computer programs to solve:

(a) $\dfrac{d^2y}{dt^2} + 2y = 5 \sin \omega t$

(b) $\dfrac{d^3x}{dt^3} - 3\dfrac{d^2x}{dt^2} + 2\dfrac{dx}{dt} + x = 0$

29. Draw an efficient computer program to solve:

$$2\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + y = 2$$

provided $y = 3$ and $dy/dt = 1$ at $t = 0$.

30. Write the equation whose analog is shown in Fig. 19.35, (C in μF, R in MΩ).

31. (a) Repeat Exercise 30 for Fig. 19.36.
(b) What is $y(t)$?

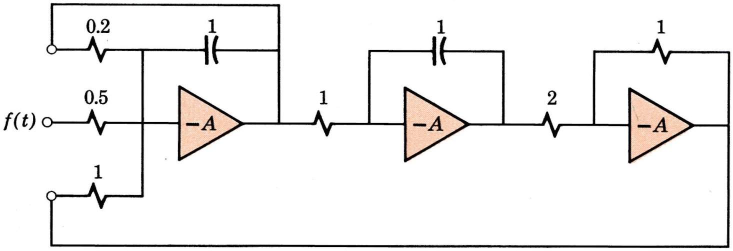


Figure 19.35

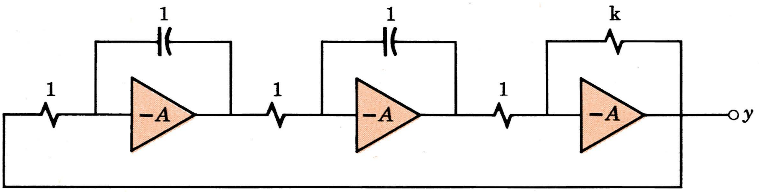


Figure 19.36

PROBLEMS

1. Working from Fig. 19.2a, derive an expression for v_i in terms of v_1 and v_o. Following the approach used in Example 7, Chapter 18, represent this equation by a block diagram with a feedback factor $R_1/(R_1 + R_F)$. Using block diagram algebra, derive an expression for gain with feedback and compare with Eq. 19-3.
2. Transform the circuit of Fig. 19.2b using the Miller technique and derive Eq. 19-3.
3. Derive a general expression for the output resistance of an op amp connected in the inverting circuit.
4. Derive a general expression for the input resistance of an op amp in the noninverting circuit.
5. An inverting amplifier ($R_1 = 1$ kΩ and $R_F = 10$ kΩ) employs an op amp with an input offset voltage of 4 mV. The offset voltage is to be balanced out by introducing a null current I_N at the junction of R_1 and R_F in Fig. 19.22.
   (a) Determine the appropriate value of I_N.
   (b) If I_N is to be obtained from a 1-MΩ potentiometer across a ± 15 V power supply, determine the setting of the potentiometer.
6. Devise a computer program to solve the equation

$$\frac{d^2y}{dt^2} + \omega^2 y = 0$$

and specify the initial conditions so that the function generated is $y = A\cos\omega t$.

7. Devise a function generator whose output is $y = 10e^{-t/5}$. Use an analog computer, programing the computer to solve an equation whose solution is the desired function. Draw a block diagram indicating the necessary input and initial conditions.

PART 3

ELECTROMECHANICAL DEVICES

Chapter 20
Energy Conversion Phenomena

Energy Sources
Energy Converters

In one sense civilization is the process whereby Man strives to become free of the restrictions imposed by the physical world. Our success in overcoming natural limitations and thereby gaining control of our environment is directly related to our ability to control small and large amounts of energy. As one example, the capabilities of our senses have been greatly extended by precisely controlled devices that enable us at one extreme to communicate over vast distances and at the other to look into the nucleus of an atom. In Part II of this book we were concerned primarily with electronic devices for controlling and processing information; consideration of the associated energy was secondary.

The civilization process has also been dependent upon Man's ability to work far beyond the limitations of our own muscles. Long ago Man learned to harness animals to supply power, and then we succeeded in converting the power of the winds and rivers to our purposes. Only recently have we learned to use the energy stored in fuels to develop mechanical power and, even more recently, electrical power. Part III is concerned primarily with devices for converting energy to and from electrical form. For practical reasons, the emphasis is on devices that employ magnetic fields in electrical–mechanical energy conversion. First, however, we take a general look at energy sources, conversion methods, and some interesting new devices that hold great promise for the future.

Energy Supply

Our interest in energy conversion methods is heightened by the population explosion, the aspiration revolution, and the politicization of petroleum. Currently about 94% of our energy comes from fossil fuels (coal 18%, natural gas 32%, and petroleum 44%) with the remainder from hydro (4%) and nuclear (2%) plants. Estimates of the world resources of fossil fuels vary widely, but it appears that at our *present* rate of consumption the supply might last a thousand years or so. The time of resource exhaustion seems comfortingly remote until we consider the changes taking place in population and standards of living.

The world population is now about 3 billion and doubling every 30 years. The per capita use of energy in the United States has doubled in the last 20 years. But if the world's peoples are to enjoy the standard of living that we accept as a matter of course, their energy consumption must increase many fold. Taking into account the unprecedented demands of an expanding population, the exhaustion of our supply of fossil fuels appears to be only a few decades away. New sources and better conversion methods must be found, and solution of the challenging problems associated with these new developments will require the concerted efforts of many engineers.

Energy Sources

The important present and future sources of energy include solar radiation, chemical reaction, and nuclear reaction; geothermal energy is an interesting speculation.

Solar. Each day there comes to the Earth from the Sun in the form of direct solar radiation an amount of energy equivalent to many years of consumption at our present rate. Summer radiation intensity is of the order of 1 kilowatt on each square meter; if efficient conversion methods were available, a small fraction of the Earth's surface could supply our current needs. Some use is made of direct radiation in powering remote installations and space laboratories, and giant power-generating satellites have been proposed, but at present the indirect effects of solar radiation are more significant.

Some of the radiation causes evaporation of sea water, which returns to the Earth as rain; a hydroelectric plant in the mountains receives its energy from the Sun indirectly. Another part of the radiation causes the growth of plants and animals that are consumed by man as fuel for his muscles. The fossil fuels represent the cumulative effect of millions of years of irradiation of living organisms and their subsequent transformation into coal and petroleum. One interesting future possibility is to grow crops that efficiently produce glucose that can be converted to alcohol, a versatile fuel.

Chemical. The basic idea of chemistry is that elements such as hydrogen, carbon, and oxygen can combine to form compounds, such as methane or carbon dioxide or water, which have properties entirely different from those of their constituents. Chemical binding is fundamentally an electrical phenomenon. In *ionic binding* one or more electrons are transferred from one atom to another, and the resulting charged ions are drawn together by the force of electrical attraction. For example, common salt consists of Na^+ ions surrounded by Cl^- ions. In *covalent binding* each atom contributes an electron to form a shared pair, and there is an increased density of negative charge in the region between the positively charged nuclei. For example, carbon forms four covalent bonds in most of its compounds.

In chemical reactions, energy is either absorbed or evolved. In charging a storage battery, the rearrangement of the reacting compounds requires an energy input equal to the increased chemical energy. In the combustion of gasoline the new compounds possess less chemical energy and the difference is available as heat. Most of the world's power now comes from the combustion of fossil fuels of which the supply is strictly limited.

Nuclear. When a molecule of hydrogen or carbon is burned, the energy released is a few electron-volts. In contrast, the energy released in nuclear *fission* is many million electron-volts per atom. In terms of energy released per unit mass, uranium fission "fuel" is nearly 3 million times as effective as coal; in terms of energy per unit cost, uranium fuel is about 400 times as effective. The total energy stored in fissionable material in the earth's crust appears to be of the same order of magnitude as the supply of fossil fuel. The number of nuclear power plants has increased rapidly, and it is predicted that by the year 2000 nearly a third of the world's power production will be from nuclear reactors.

When heavy nuclei split in the process of fission, about 0.1% of the original mass is converted into energy. In the *fusion* of two deuterons (isotopes of hydrogen) to form a helium atom, about 0.6% of the original mass is converted into energy. Since an appreciable fraction (0.015%) of all hydrogen is in the form of deuterium, ordinary sea water represents an unlimited supply of energy if fusion reactions can be initiated and controlled. To overcome the electrostatic repulsion between two deuterons requires energies corresponding to temperatures of millions of degrees. Such *thermonuclear* reactions can be initiated by the explosion of an A-bomb (nuclear fission) and sustained by the energy released by fusion; the result is an H-bomb that may yield energies equivalent to many millions of tons of TNT. If useful power is to be obtained from fusion, ways must be found to confine and control the thermonuclear reactions. A promising approach is to use powerful laser pulses to heat a tiny fuel pellet, initiating fusion and producing energy before it can fly apart.

Geothermal. The heat of the earth's interior is a potential energy source. Natural steam or steam produced by circulating water through fractured rocks can contribute significantly to our energy resources if the difficult environmental and technical problems (noise and air pollution, corrosion) can be solved.

Energy and Entropy

In a general consideration of energy conversion there are two fundamental principles that were originally formulated in connection with thermodynamics. The *first law* of thermodynamics states that energy can be converted from one form to another, but it can be neither created nor destroyed. In applying this so-called "law of conservation of energy," matter is included as a form of energy. One application is the *heat balance,* a detailed expression of the fact that all the energy inputs to a system must equal the sum of all the energy outputs plus any increase in energy storage. We shall make use of this approach in analyzing the performance of electromechanical devices.

The *second law* of thermodynamics says that no device, actual or ideal, can both continuously and completely convert heat into work; some of the heat is unavailable and must be rejected. For example, no engine could extract heat from sea water and convert it to work with no effect other than the cooling of the water. This so-called "law of degradation of energy" stipulates what transformations of energy are possible, whereas the first law governs the energy relations in a possible transformation. The unavailability of energy is measured by a property called *entropy,* which has some of the characteristics of probability; a uniform distribution of energy corresponds to a high entropy. We say that the energy of the universe remains constant, while the entropy tends toward a maximum. One of the primary activities of the engineer is directing the inevitable degradation of energy so that in the process some useful result is obtained.

An important characteristic of energy is whether it has an *ordered* or *disordered* form. The energy of an electric current in an inductor or the energy of a rotating flywheel is said to be ordered. In contrast, the random thermal energy of electrons or atoms in a solid is said to be disordered. Energy conversion from an ordered form to a disordered form, as in heating by an electric current or friction braking of a flywheel, can be achieved with 100% efficiency. The efficiency of energy conversion between ordered forms, as in an electric generator, can approach 100%. However, the efficiency of conversion from disordered to ordered form, as in a steam turbine (thermal to mechanical), is limited by the requirement that some energy be rejected; the maximum conceivable efficiency of such a conversion is $(T_1 - T_2)/T_1$ where T_1 is the highest temperature (K) in the cycle and T_2 is the lowest naturally available temperature. This concept of a limiting efficiency for heat engines was presented in 1824 by Sadi Carnot, a 28-year old French engineer.

Electrical Energy Generation

One reason for the usefulness of electrical energy is that it is highly ordered and therefore it can be converted efficiently into other forms. On the other hand, it cannot be stored conveniently, so it is usually generated as it is needed. The great variety of energy conversion possibilities is displayed in Fig. 20.1.

Most electrical energy is still generated in the process used by Edison in his first central station in 1882. Fossil fuel is burned to obtain heat to convert water into steam to drive mechanical engines to force electrical conductors through a magnetic field to generate voltage and current. An atomic power plant differs only in that the heat is obtained from nuclear reaction. In a hydroelectric plant the electrical generator is driven by a water turbine instead of a steam turbine.

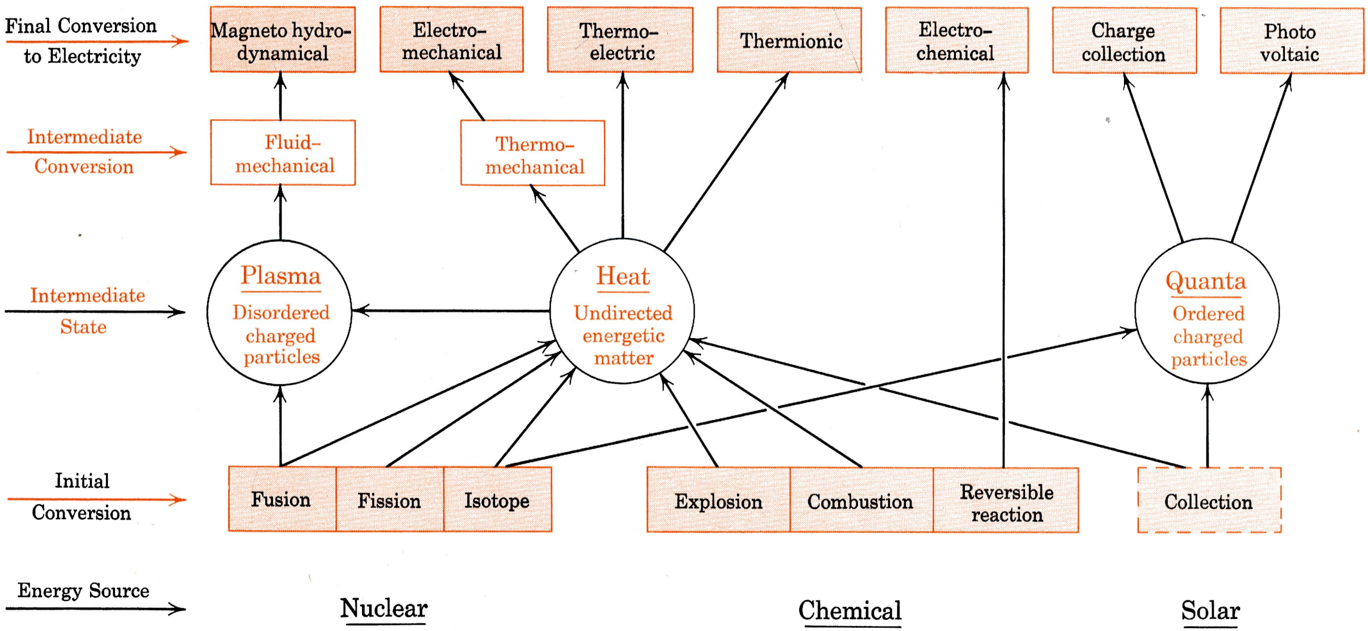

Final Conversion to Electricity
Intermediate Conversion
Intermediate State
Initial Conversion
Energy Source
Magneto hydro-dynamical
Electro-mechanical
Thermo-electric
Thermionic
Electro-chemical
Charge collection
Photo voltaic
Fluid-mechanical
Thermo-mechanical
Plasma
Disordered charged particles
Heat
Undirected energetic matter
Quanta
Ordered charged particles
Fusion
Fission
Isotope
Explosion
Combustion
Reversible reaction
Collection
Nuclear
Chemical
Solar

Figure 20.1 Energy conversion to electrical form. (Courtesy General Electric Company.)

The virtue of magnetic-field generators is that high permeability materials permit extremely high energy densities and therefore high output from a machine of given volume. The low overall efficiency of the thermal–mechanical–electrical generating system (typically 40% in 1974, increased from < 20% in 1920) has been tolerated in the past because of the abundance of low-cost fuel.

Since chemical energy is more ordered than thermal energy, higher conversion efficiencies are to be expected in generators employing chemical reaction. In a *voltaic cell,* two dissimilar electrodes are separated by an electrolyte in which conduction takes place by the motion of positive and negative ions. The chemical reactions at the electrode–electrolyte surfaces provide the energy for continuous current production. In a *primary* cell, such as the ordinary dry cell or mercury battery, the energy conversion is accompanied by irreversible changes in cell composition and the cell has a limited life. In a *secondary* cell, such as the lead–acid storage battery, the chemical reactions are reversible. During discharge, lead sulphate and water are formed; the water dilutes the sulfuric acid electrolyte and the specific gravity of the electrolyte is an indication of the state of charge. The battery is charged by sending a current from an external source through the electrolyte in the opposite direction. The reactions are reversed, lead and lead dioxide are formed at the negative and positive plates, respectively, and the battery is restored to its original condition.

Fuel Cells

A device that continuously converts the chemical energy in a fuel directly into electrical energy is very attractive. By avoiding the intermediate thermal energy stage, the cost and complexity of thermal-to-mechanical-to-electrical conversion apparatus is eliminated and the Carnot efficiency limitation is removed. The *fuel cell* is such a device; already it has been used to supply power in space vehicles and it shows great promise for the future.

The operation of a hydrogen–oxygen fuel cell of the type used in the Gemini spacecraft is illustrated in Fig. 20.2. The cell consists of two chambers and two

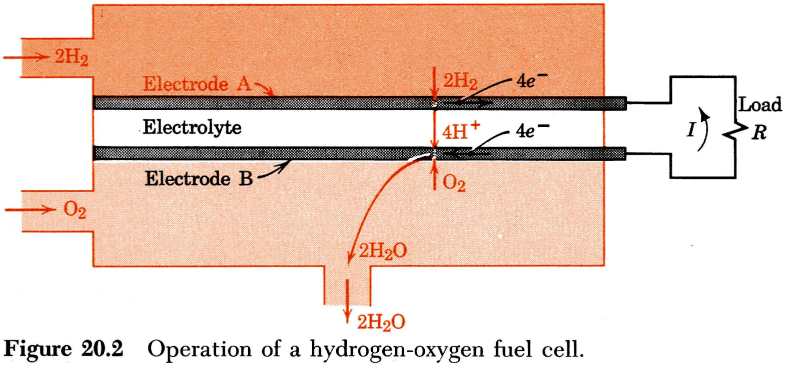


Figure 20.2 Operation of a hydrogen-oxygen fuel cell.

porous electrodes separated by an electrolyte. Hydrogen supplied to the upper chamber diffuses through electrode *A* and, in the presence of a catalyst, reacts with the electrolyte to form positive ions and free electrons. The ions migrate through the electrolyte to electrode B. The electrons pass through the external load circuit and then to electrode B where they combine with oxygen and the ions to form water. The reaction at electrode A is

$$2H_2 \rightarrow 4H^+ + 4e^-$$

and that at electrode B is

$$4H^+ + 4e^- + O_2 \rightarrow 2H_2O$$

The important point is that the electrons are forced to do useful work before the reaction is completed. In a properly designed cell, most of the energy that would appear as heat in a combustion reaction is available as electrical energy.

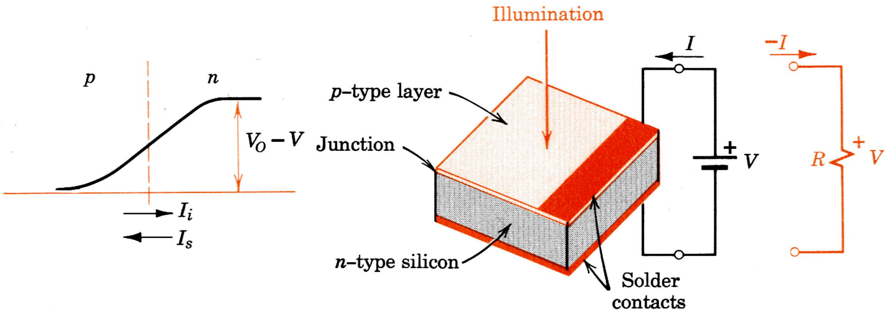
A "hydrox" fuel battery gives several times as much energy as the same weight of storage battery, and the pint of drinking water produced with each kilowatt-hour is a useful by-product. Theoretical efficiencies of over 90% are indicated for some types of fuel cells and efficiencies over 80% have been obtained in the laboratory. Fuel cells that burn ordinary hydrocarbon fuels in air are already available. Inexpensive hydrocarbon fuel batteries operating at 80% efficiency could double the energy potential of our fossil fuel resources. High efficiency at partial load and quiet operation without exhaust are additional advantages.

Solar Cells

Knowing the tremendous energy available in the form of solar radiation, we can appreciate the possibilities of a device that converts light energy directly into electrical energy. The semiconductor *solar cell* operates with fair efficiency, has an unlimited life, and has a high power capacity per unit weight. Already it is an important source of power for long-term satellites, and continued improvement in operating characteristics will make it competitive in other applications.

The operation can be explained in terms of our knowledge of *pn* junction diodes. As a result of the diffusion of majority carriers near the junction (see Figs. 11.10 and 11.11) a potential barrier is created. Under conditions of forward bias (Fig. 20.3a), the potential barrier is slightly reduced, the number of majority carriers with sufficient energy to climb the potential "hill" is increased, and a net forward current flows. If now the junction region is irradiated with photons possessing sufficient energy, electron-hole pairs are created and the density of minority carriers is greatly increased. (On a percentage basis, the increase in majority carriers is much smaller.) Practically all the minority carriers drift across the junction and contribute a component I_P to the diode current. Equation 11-21 becomes

$$I = I_s e^{eV/kT} - I_s - I_P \tag{20-1}$$



(a) Potential hill (b) Construction and connection

Figure 20.3 Solar cell principle and construction.

since I_P is in the same direction as the I_s due to thermally generated carriers. For high light intensity and low bias voltage the net current

$$I = I_s(e^{eV/kT} - 1) - I_P \tag{20-2}$$

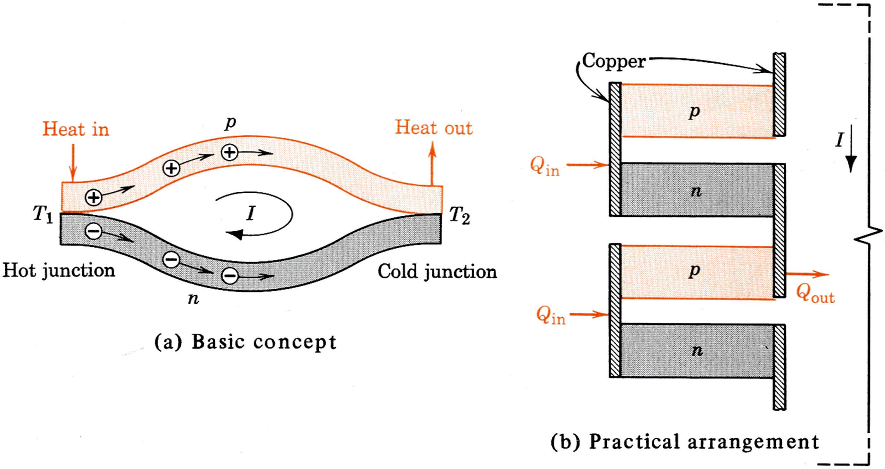
is negative and the solar cell is *charging* the bias battery. Such a *pn* junction diode is a source of electrical energy. Alternatively, the solar cell may supply a load R where $V = (-I)R$.

A practical solar cell is constructed so that the junction is exposed to the light. As shown in Fig. 20.3b, a very thin layer of *p*-type material is created by diffusion of acceptor atoms into a heavily doped *n*-type silicon wafer. Electrical contact is provided by a thin translucent nickel plating (not shown) and a solder contact over the bottom and along one edge of the top. Because the current and voltage output of a single cell are small, cells are usually connected in series and parallel to form solar batteries. Overall efficiencies are in the range from 10 to 15% and satellite units provide 10 to 15 watts per pound. These favorable operating characteristics, the plentiful supply of the chief ingredient silicon, and the ease of mass production of the units all point to increased use of solar cells for energy conversion.

Thermoelectric Converters

The fuel cell is based on an old idea; it has become important recently because of advances in basic science that resulted in higher efficiencies. The solar cell is a modern development based on new insights into semiconductor phenomena. The *thermoelectric* converter represents a new application of an old idea. In 1821, Seebeck noted that heat applied to a junction of dissimilar metals could produce a small current in a closed circuit. The efficiency of conversion was so low, however, that his *thermocouples* were applied only for measurement purposes. Advances in solid-state theory and technology have led to higher efficiencies and the design of practical energy converters.

The operation of a semiconductor converter is similar to that of a solar cell in that energy is added to generate electron-hole pairs near a *pn* junction; the directional characteristic of the junction produces a separation of charge and a net current flow that is available to do work. As shown in Fig. 20.4a, the hot junction is maintained at temperature T_1 by an input of thermal energy. Holes generated near the junction drift across the junction into the *p* region, and electrons drift into the *n* region. At the cold junction, the effect is less and therefore there is a net flow of current consisting of electrons in the *n* material and holes in *p* material.



(a) Basic concept

(b) Practical arrangement

Figure 20.4 Semiconductor thermoelectric converters.

Note that, if minority carriers drift across a *pn* junction, they must gain energy and this energy is provided by the heat source. If minority carriers are *forced* through junction 1 in the direction shown in Fig. 20.4a, junction 1 is *cooled.* This is the *Peltier effect* and permits the use of a *pn* junction as a heater or a cooler, depending on the direction of current flow. In this case, electrical energy is used to provide a *heat pump* effect.

Figure 20.4b shows a more practical form of converter arrangement. Because good thermoelectric materials are usually poor thermal conductors, the active elements are kept short and the heat input is by way of a good thermal and electrical conductor such as copper. The net thermoelectric effect is unaffected. Since the output is only about 100 μV per K, many junctions are connected in series; in terms of heat flow, the elements are in parallel.

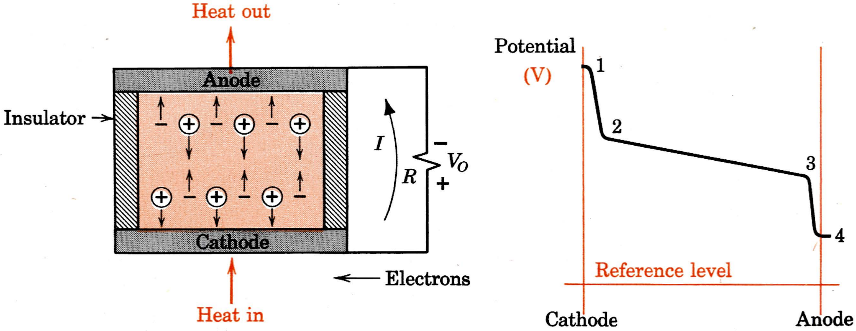
A great advantage of the thermoelectric converter is that any form of thermal energy can be used. However, since the input energy is disordered, the efficiency of this device can only approach the Carnot ideal $(T_1 - T_2)/T_1$. Materials have been developed that provide conversion efficiencies as high as 13%, but practical units operate at considerably lower efficiencies. Sacrificing efficiency for simplicity, Russian engineers have developed a kerosene-burning unit that supplies the few watts of power needed for a radio receiver in a remote location.

Thermionic Converters

Just as the semiconductor diode led to the modern thermoelectric converter, the vacuum diode is the basis for another new device for the direct conversion of heat to electrical energy. In the *thermionic* converter, heat supplied to the cathode (emitter) provides the electrons with more than enough energy to overcome the cathode work function. The electrons escape with sufficient energy to move to the anode (collector) against a potential difference that includes an output voltage across a load. In a practical converter, the retarding effect of the space charge must be kept low and the anode temperature must be kept low to prevent anode emission.

The space-charge effect can be reduced by making the cathode–anode spacing very small; this is difficult to do in large units. Another possibility is to neutralize the negative space charge with slow-moving positive ions. Filling the space with cesium vapor greatly reduces the space charge and also improves the effective work functions of the cathode and anode surfaces. The operation of a vapor-filled thermionic converter is shown diagrammatically in Fig. 20.5a.† The heat supplied to the cathode produces a great quantity of electrons with energies corresponding to the potential at point 1 in Fig. 20.5b. Just outside the emitter surface negative space charge is high and an appreciable part of the escape energy is lost. A much smaller potential drop (from V_2 to V_3) occurs across the balance of the neutralized cathode–anode space. A free electron just outside the anode gives up energy corresponding to $V_3 - V_4$ upon entering the anode lattice structure. The energy available for useful work in an external load corresponds to potential V_4.

The thermionic converter is a high-temperature, low-voltage, high-current



(a) Construction (b) Potential variation

Figure 20.5 A vapor-filled thermionic converter.

---

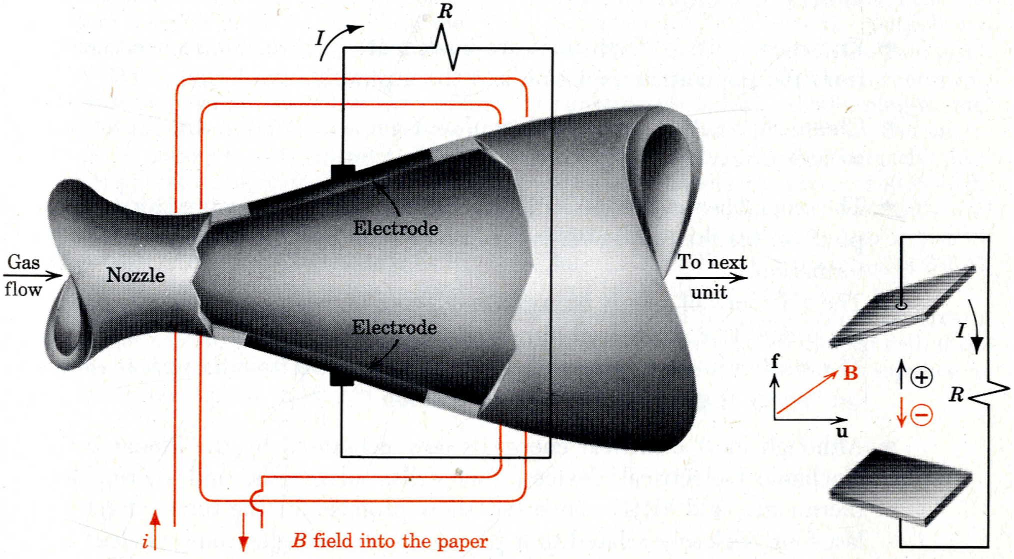
† See V. C. Wilson, "Thermionic Power Generation," *IEEE Spectrum*, May 1964, p. 75.

device. Cathodes operate efficiently at temperatures around 2000 K and this matches the delivery temperature of energy from a solar furnace or a nuclear reactor. Efficiencies are limited by the second law of thermodynamics but are expected soon to be near 20%. Since the unit is light and rugged, it may be desirable for compact portable power plants. Another promising application is as the first stage in a nuclear power plant; the heat rejected at the anode could be used by a conventional steam turbine and the overall plant efficiency would be increased.

Magnetohydrodynamic Converters

Creation involves the recognition of new relations among known facts. Consider the following facts: A voltage can be induced in a conductor moving in a magnetic field; thermal energy can be converted to kinetic energy in the form of a high-velocity stream of gas; gases can be made conducting by ionization or by seeding with ionized vapor. Looking at these facts in the light of knowledge gained in studying the behavior of gas tubes, the propagation of shock waves, and air ionization during missile reentry, engineers conceived a generator-engine based on the new science of *magnetohydrodynamics* (MHD). In the next decade devices employing this concept may be propelling vehicles in the outer reaches of the solar system or producing huge quantities of electrical energy here on Earth.

Although the design of practical MHD generators is challenging some of the best engineers and scientists, the principle of operation is simple. As shown in Fig. 20.6, ionized gas or *plasma* leaves the nozzle at high velocity and enters



(a) Construction (b) Field relations

Figure 20.6 An elementary magnetohydrodynamic generator.

a transverse magnetic field. The charged particles experience a force $\mathbf{f} = q\mathbf{u} \times \mathbf{B}$ that drives positive ions upward and negative ions or electrons downward. If a conducting path is provided, an electrical current flows from the upper electrode through an external load to the lower electrode. The electrical energy is gained at the expense of mechanical energy abstracted from the moving gas just as in a turbine. The energy conversion process is reversible; if electrical energy is supplied to force an electric current in the opposite direction, the average velocity of the plasma is increased and the device is operating as an engine.

The operating characteristics of an MHD generator provide a guide to its potential applications. It works well at high temperatures (around 3000 K) and this indicates the possibility of high overall efficiency (up to 55%) in a power plant. The output increases directly as the volume increases, whereas the losses increase more slowly; MHD would be more competitive in large units (several hundred megawatts). The demands on channel and electrode materials are severe under continuous operation; however, a solid-propellant rocket motor and a simple MHD converter could supply large amounts of electrical power in short bursts. In spite of the unsolved problems in obtaining high conductivity plasmas, high combustion and heat transfer rates, and durable materials, MHD power generation appears to have an important future.

SUMMARY

- The future course of civilization depends on Man's ability to improve the efficiency with which existing energy resources are used and to develop new sources of energy.
- Known supplies of fossil fuels are inadequate to meet future needs resulting from the population explosion and the aspiration revolution.
- Chemical reaction, solar radiation, and nuclear reaction are the important sources of energy for the near and distant future.
- The second law of thermodynamics indicates what energy transformations are possible; the first law governs the energy relations in a possible energy transformation.
  The efficiency of energy conversion depends on whether the energy quantities are ordered or disordered (random).
  The maximum conceivable efficiency of conversion from thermal to mechanical energy is given by the Carnot relation.
- Although most electrical energy is now generated by traditional thermal–mechanical–electrical devices, fuel cells, solar cells, and thermoelectric, thermionic, and MHD converters show promise for the future. Each of the last four is closely related to a previously studied electronic device.

PROBLEMS

The following problems are similar to practical engineering assignments in that the information available is inadequate for precise calculations. Simplifying assumptions must be made; your instructor may indicate appropriate values or handbooks can be consulted. Be sure to state all assumptions made and to indicate clearly the line of reasoning followed.

1. Compare the energy to be used in the next two decades to the total energy used up to this time.
2. The heating value of a pound of coal is about 10^7 J. Estimate the amount of coal required to light a 100-W bulb 6 hr/day for a year.
3. The major home appliances and some "typical" energy consumptions are: water heater 15 kWh/day, air-conditioner 2 kW, refrigerator-freezer 4 kWh/day, dishwasher 0.6 kWh/cycle, range and oven 3 kWh/day, washer and dryer 3 kWh/cycle, and color TV 200 W. Estimate the annual energy cost for a home in your community.
4. It has been proposed that gas water heaters be banned to reduce air pollution. Compare the natural gas consumed in a home water heater (60% efficiency) with that consumed at a distant power plant to produce electricity to transmit to the home to heat the same amount of water. Draw a conclusion.
5. A 1000-MW power plant uses river water to absorb the energy rejected in the condenser of the steam turbine. Estimate the gallons per minute of water required if the water temperature increase ("thermal pollution") is to be less than 15°F.
6. Compare a 20-gallon gasoline tank and a storage battery occupying the same volume on each of the following criteria: total energy content, mechanical energy available for automotive propulsion, and energy cost per mile at in-town speeds.
7. In a large power plant the steam input to the turbine is at 1000°F and the condenser-cooling water from a lake is at 50°F. The overall plant efficiency is 40%. If improved engineering would permit using steam at 1100°F, what overall efficiency would be expected?
8. If the steam turbine of Problem 7 is preceded by a thermionic converter with a cathode at 3000°F and anode at 2000°F, what improvements in overall efficiency might be expected?
9. If the steam turbine of Problem 7 is preceded by an MHD generator operating at 4000°F at the intake and 2500°F at the exit with dc-to-ac conversion 95% efficient, what improvement in overall efficiency might be expected?
10. In a proposed plant, a 1000-MW nuclear reactor delivers compressed helium gas at 1400°C to an MHD generator that exhausts at 1050°C. (The plasma ionizing system and superconducting magnet require 2.5% of the MHD electrical output, and the dc-ac conversion is 97.5% efficient.) The MHD exhaust gas then generates steam to drive a turbine. The turbine provides 200 MW to compress the helium gas and rejects 500 MW; the remainder drives a 98% efficient ac generator. Estimate the output and overall efficiency of this MHD-steam plant.
11. A fuel cell weighing 25 lb has a continuous capacity of 200 W; a 6-lb plug-in fuel cartridge provides 15 hr of full-load operation. Also available are rechargeable batteries with a specific capacity of 40 W-h/lb. Compare the weights of fuel cells and batteries to be transported to a remote location for 10 hours of operation at 200 W and for 10 days of operation.
12. Estimate the fraction of the total area of your city to be covered by solar cells if the total electrical load were to be supplied by solar radiation.

Chapter 21

Magnetic Fields and Circuits

Magnetic Fields
Magnetic Circuits

As pointed out in Chapter 20, there are some exciting new developments in energy conversion, but at present practically all electrical power is *generated* by devices employing magnetic fields. Also, most electrical energy is *consumed* in devices employing magnetic fields in the conversion of electrical to mechanical energy. Most electromechanical devices could more properly be called *electromagnetomechanical* devices because magnetic fields provide the essential coupling in the energy-conversion process. The predominance of magnetic coupling is due to the high energy densities obtainable with commonly available magnetic materials; high energy density results in high power capacity per unit volume of machine.

We have already worked with magnetic fields in two different situations. We defined inductance as a measure of the ability of a circuit component to store energy in a magnetic field (Chapter 1) and employed the resulting v-i characteristic in analyzing circuits exhibiting this property. Also, we defined magnetic flux density in terms of the force on a moving charge and used this concept in studying the motion of electrons in a uniform magnetic field (Chapter 10). In the next three chapters we are going to study the operating principles of some important energy conversion devices in order to learn to predict their performance. We need a quantitative understanding of magnetic fields and the magnetic circuits employed to establish them.

Magnetic Fields

The behavior of an electric circuit can be completely described in terms of the voltage and current at various points along the path constituting the circuit. In contrast, it is characteristic of fields that they are distributed throughout a region and must be defined in terms of two or three dimensions. What is a magnetic field? About all we can say is that it is a region of space with some very useful properties. Does it really exist? From our standpoint it is a convenient concept for describing and predicting the behavior of devices that do exist.

Magnetic Flux and Flux Density

Magnetic fields are created by electric charge in motion and, in turn, the strength of magnetic fields is measured by the force exerted on a moving charge. In vector notation, the defining equation (Eq. 1-6) is

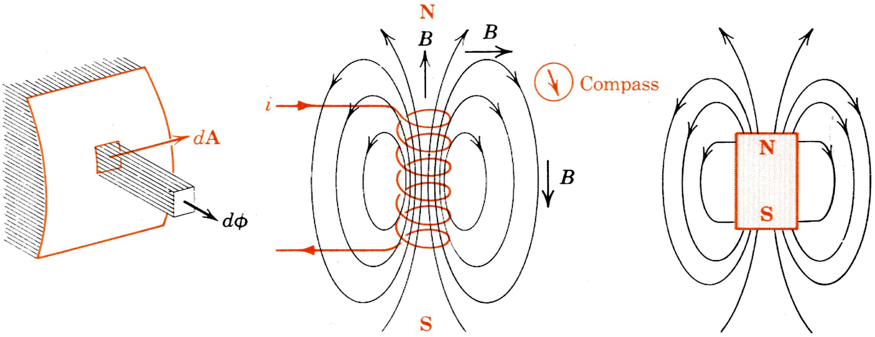
$$\mathbf{f} = q\mathbf{u} \times \mathbf{B} \tag{21-1}$$

A *magnetic flux density* B of one tesla[†] "exists" when a charge q of 1 coulomb moving normal to the field with a velocity u of 1 meter per second experiences a force f of 1 newton. If $\mathbf{u}$ is at an angle θ with respect to $\mathbf{B}$ (see Fig. 10.3), the direction of $\mathbf{f}$ is normal to the plane containing $\mathbf{u}$ and $\mathbf{B}$ and the magnitude of $\mathbf{f}$ is $quB \sin\theta$.

The summation obtained by integrating flux density over an area (Fig. 21.1a) is *magnetic flux* ϕ in webers defined by

$$\phi = \int \mathbf{B} \cdot d\mathbf{A} \tag{21-2}$$

In describing a magnetic field, we represent magnetic flux by *lines of magnetic force* or just *lines*. The lines are drawn tangent to the flux density vector at any point in the field (Fig. 21.1b). The lines are close together where the flux density is high. An important fact is that the amount of magnetic flux leaving any closed surface is just equal to the amount entering; in other words, magnetic flux lines are continuous. This fact can be expressed mathematically by the equation



(a) Element of flux (b) Field of a solenoid coil (c) Field of a magnet

Figure 21.1 Two-dimensional representation of magnetic fields.

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0 \tag{21-3}$$

[†] For conversion from MKS units to CGS units, see Appendix.

where the symbol $\oint$ indicates integration over a closed surface. In Fig. 21.1c the lines close within the magnet.

Fields Due to Currents

In establishing magnetic fields, we are interested in arranging the motion of charge to achieve a maximum effect. Copper conductors wound in compact coils provide an effective arrangement in many situations. To determine the magnetic effect of charges moving in conductors in various configurations, we first consider the effect of a current i flowing in a short element of conductor ds (Fig. 21.2). If all the charge dq in the element moves a distance ds in time dt, the velocity u is ds/dt or $ds = u\,dt$. Then

$$i\,ds = \frac{dq}{dt}(u\,dt) = dq\,u \qquad (21\text{-}4)$$

and we can see that current i in element ds is equivalent to charge dq at velocity u.

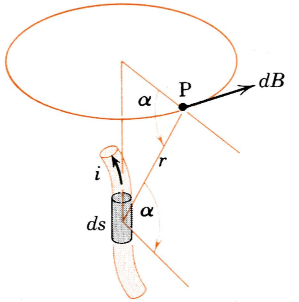


Figure 21.2 Magnetic field due to a current.

Experiments (first performed by the ingenious French physicist André Ampère in 1820) indicate that the contribution of current i in element ds to the magnetic flux density dB at point P in a homogeneous material is proportional to the current i and the cosine of angle α and inversely proportional to the square of the distance r. In MKS units,

$$dB = \mu\frac{i\,ds\cos\alpha}{4\pi r^2} \qquad (21\text{-}5)$$

The factor μ is a property of the material surrounding the conductor and is called the *permeability* in webers/ampere-meter or the equivalent henrys/meter. The direction of the dB vector is tangent to the circle with its center on the extension of ds and passing through the point P as shown. A convenient rule is:

> *If the conductor is grasped in the right hand with the thumb extending in the direction of i, the fingers curl in the direction of B.*

Magnetic Field Intensity

The permeability of free space is $\mu_o = 4\pi \times 10^{-7}$ henrys per meter; the ratio of the permeability of any substance to that of free space is called the *relative permeability* μ_r, a dimensionless number. The relative permeability for most materials is near unity; the permeabilities of air and copper, for example, are practically the same as that of free space. The relative permeabilities of the *ferromagnetic* materials (iron, cobalt, nickel, and their alloys) may be in the hundreds of thousands. In other words, the magnetic flux density produced by a given current in a coil wound on a ferromagnetic core may be several thousand times as great as the flux density produced in air by the same current and coil.

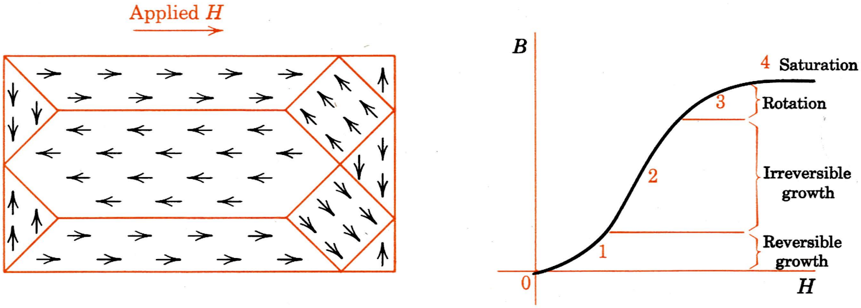
To eliminate the effect of the medium, it is convenient to define the *magnetic field intensity H* where

$$\mathbf{H} = \frac{\mathbf{B}}{\mu} \tag{21-6}$$

in amperes per meter. The magnetic field intensity is a measure of the *tendency* of a moving charge to produce flux density; the actual value of B produced depends on the permeability of the medium.

Ferromagnetism

The spin of an orbital electron constitutes charge in motion, and therefore magnetic effects occur on an atomic level. Since in the atoms of most materials these electron spins are cancelled out by other electron spins, there is no net effect. In ferromagnetic materials, however, there are unbalanced electron spins and also a tendency for neighboring atoms to align themselves, so that their magnetic effects all add up.† In a specimen of unmagnetized ferromagnetic material, there are small *domains* in which all the atoms are aligned. Each domain (on the order of a thousandth of an inch in extent) is a region of intense magnetization. However, the domains are randomly oriented (Fig. 21.3a) and the specimen exhibits no net external magnetic field.



(a) Domains in an unmagnetized specimen **(b) Magnetization curve**

Figure 21.3 Magnetization of an iron specimen.

If an external field is applied, there is a tendency for the tiny magnets to align with the applied magnetic field or *polarize* just as a compass needle tends to align itself with the Earth's field. At low values of field intensity H (region 1 in Fig. 21.3b), domains nearly aligned with the applied field grow at the expense of adjacent, less favorably aligned domains in an elastic (reversible) process. This results in an increase in flux density B over that expected in free space. As H is increased (region 2), the direction of magnetization of misaligned domains switches, in an irreversible process, and this contributes to a rapid increase in B. At higher values of H (region 3), the directions of magnetization rotate until the contributions of all domains are aligned with the applied field. A further increase in field intensity produces no further effect within the ferromagnetic material and the material is said to be *saturated* (region 4). Commercial magnetic steels (usually called "iron") tend to saturate at flux densities of 1 to 2 teslas. For very high values of H the slope of the B–H curve approaches μ_o.

---

† See J. M. Ham and G. R. Slemon, *Scientific Basis of Electrical Engineering*, John Wiley & Sons, New York, 1961.

The flux density in a ferromagnetic material is the sum of the effects due to the applied field intensity H and the *magnetic polarization* M produced within the material. This relation can be expressed by the equation

$$B = \mu_o(H + M) \tag{21-7}$$

This can be rewritten as

$$B = \mu_o\left(1 + \frac{M}{H}\right)H = \mu_o\mu_r H = \mu H \tag{21-8}$$

From Fig. 21.3b it is clear that M/H is not a constant and therefore μ_r, the relative permeability, is not a constant. Because μ_r is greatly affected by rolling, stamping, and other material processing and because it depends on the previous history of magnetization (see Fig. 22.4), μ_r is seldom known precisely. For many calculations an average, constant value of μ_r can be assumed.

Field Around a Long Straight Conductor

Equation 21-5 expresses the contribution dB at point P by a current i in element ds. To determine the field around a conductor carrying current, we can obtain the total effect by integration. For the infinitely long straight conductor of Fig. 21.4, $r = D/\cos\alpha$. Then, considering differentials,

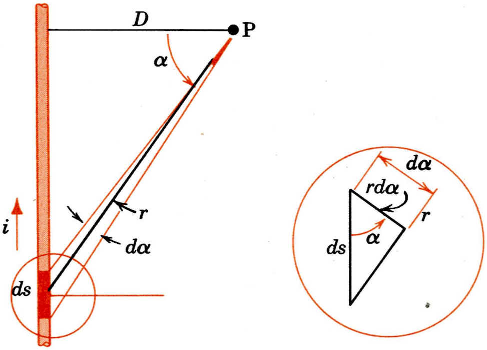


Figure 21.4 Field around a long conductor.

$$ds \cos\alpha = r\,d\alpha = \frac{D}{\cos\alpha}d\alpha$$

and

$$\frac{ds \cos\alpha}{r^2} = \frac{(D/\cos\alpha)\,d\alpha}{D^2/\cos^2\alpha} = \frac{\cos\alpha\,d\alpha}{D}$$

Substituting in Eq. 21-5 and integrating,

$$B = \frac{\mu i}{4\pi}\int_{-\infty}^{+\infty} \frac{\cos\alpha}{r^2}\,ds = \frac{\mu i}{4\pi D}\int_{-\pi/2}^{+\pi/2} \cos\alpha\, d\alpha = \frac{\mu i}{4\pi D}\Big[\sin\alpha\Big]_{-\pi/2}^{+\pi/2} = \frac{\mu i}{2\pi D} \tag{21-9}$$

In words, flux density B is directly proportional to current i and varies inversely as the distance D.

Example 1

Derive an expression for the flux density at the center of a compact N-turn circular coil.

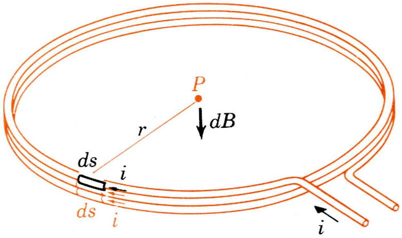


Figure 21.5 Flux density produced by a compact circular coil.

For the configuration of Fig. 21.5, $\alpha = 0$, $\cos\alpha = 1$, and Ampere's law becomes

$$dB = \frac{\mu i\, ds}{4\pi r^2}$$

The total flux density for a full turn is

$$B = \int dB = \frac{\mu i}{4\pi r^2}\int_0^{2\pi r} ds = \frac{\mu i(2\pi r)}{4\pi r^2} = \frac{\mu i}{2r}$$

For a compact N-turn coil, there are N similarly oriented elements carrying current i.

$$\therefore B = N\frac{\mu i}{4\pi r^2}\int_0^{2\pi r} ds = \frac{\mu N i}{2r} \tag{21-10}$$

An important principle can be demonstrated by considering the field intensity around a long straight conductor. From Eqs. 21-6 and 21-9,

$$H = \frac{B}{\mu} = \frac{i}{2\pi D}$$

Along a circle of radius D, field intensity is constant and the line integral is

$$\oint \mathbf{H}\cdot d\mathbf{l} = \frac{i}{2\pi D}\cdot 2\pi D = i \tag{21-11}$$

Although the relation is derived here for a special case, it has the following general interpretation.

> *The line integral of field intensity along any closed path is just equal to the current linked.*

The quantity $\int \mathbf{H}\cdot d\mathbf{l}$ is called the *magnetomotive force* (mmf) because in magnetic circuits it plays a role analogous to the electromotive force in electric circuits. This principle, sometimes called the *mmf law,* is very useful in determining the fields due to currents in conductors of various configurations.

Field Produced by a Toroidal Coil

One conductor configuration for creating a strong magnetic field consists of many turns of wire wound on a cylindrical form; such a *solenoid* (Fig. 21.1b) may have an air core or, for greater flux density, an iron core. Another common form is the *toroid* of Fig. 21.6 in which the wire is wound on a doughnut-shaped core. Let us use the mmf law to investigate the field in the vicinity of the toroidal coil.

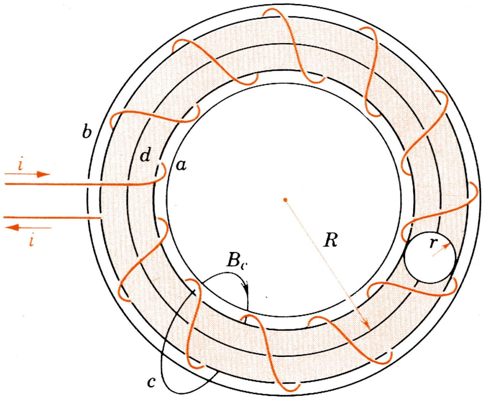


Figure 21.6 A toroidal coil.

Since closed path a links no current, the mmf along the path is zero; since $\int dl$ is finite, net H must be zero. We conclude that there is no field intensity and no magnetic flux in the direction of path a. Following the same reasoning, what conclusion do you reach regarding path b?

Closed path c links current i once and, therefore, an mmf of i A exists. The net current flow is clockwise around the toroid. Applying the righthand rule, the resulting H and $B = \mu_0 H$ are in the direction of the arrow. A small amount of magnetic flux goes through the hole and into the paper. The flux density at any point in the interior or exterior could be calculated by using Ampere's law (Eq. 21-5).

If there are N turns on the toroid, closed path d links the current i N times. We know that the magnetic field is due to charge in motion. Insofar as path d is concerned, there are N conducting paths in parallel, each carrying a current $i = q/t$. We conclude that the total effect is as if a current Ni were linked once. By Eq. 21-11 the magnetomotive force in ampere-turns† is

$$\oint \mathbf{H} \cdot d\mathbf{l} = H(2\pi R) = Ni$$

and

$$B = \mu H = \frac{\mu}{2\pi R} Ni \tag{21-12}$$

---

† Since the number of turns is a dimensionless quantity, the *unit* of mmf is the *ampere*. However, we shall use "ampere-turns" (abbreviated A-t) to emphasize the significance of the number of turns in the usual case of a multiturn coil.

If the toroid has a circular cross-section with a radius r, the core area is πr^2. If r is small compared to R, the flux density can be assumed to be uniform at the value given by Eq. 21-12 where R is the mean radius. On the basis of this assumption,

$$\phi = BA = \frac{\mu}{2\pi R} Ni(\pi r^2) = \frac{\mu r^2}{2R} Ni \tag{21-13}$$

Example 2

A coil consists of 1000 turns wound on a toroidal core with $R = 6$ cm and $r = 1$ cm. To establish a total magnetic flux of 0.2 mWb in a nonmagnetic core, what current is required? Repeat for an iron core with a relative permeability of 2000.

For a nonmagnetic core $\mu = \mu_o$. By Eq. 21-13,

$$i = \frac{2R\phi}{\mu r^2 N} = \frac{2 \times 6 \times 10^{-2} \times 2 \times 10^{-4}}{4\pi \times 10^{-7} \times 10^{-4} \times 10^{3}} \cong 190 \text{ A}$$

For an iron core, $\mu = \mu_o \mu_r = 2000\, \mu_o$, and

$$i = \frac{190}{2000} = 0.095 \text{ A} = 95 \text{ mA}$$

Note that the flux density is

$$B = \frac{\phi}{A} = \frac{2 \times 10^{-4}}{\pi \times 10^{-4}} \cong 0.64 \text{ T}$$

which is well below the saturation value for commercial "iron."

Letting $l = 2\pi R$ represent the length of the core and $\mathfrak{F} = Ni$ the effective mmf, Eq. 21-13 can be rewritten as

$$\phi = \mu \frac{A}{l} \mathfrak{F} \tag{21-14}$$

Does the form of this equation look familiar? Is there an analogy between this equation and Eq. 11-3? Do you see any possibility of considering a magnetic field as a "circuit"?

Summary of Magnetic Field Relations

Before leaving the subject of magnetic fields, let us list six important relations:

$$\mathbf{f} = q\mathbf{u} \times \mathbf{B} \qquad \int \mathbf{B} \cdot d\mathbf{A} = \phi$$

$$dB = \frac{\mu i \, ds \cos\alpha}{4\pi r^2} \qquad \oint \mathbf{B} \cdot d\mathbf{A} = 0$$

$$\mathbf{B} = \mu \mathbf{H} \qquad \oint \mathbf{H} \cdot d\mathbf{l} = Ni$$

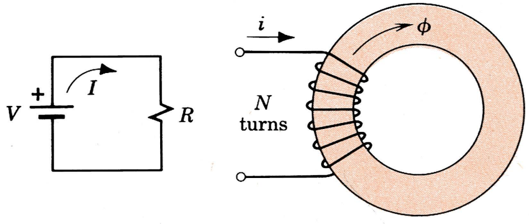
Three of these relations are definitions. Can you identify them? The other three relations are based on experimental observations. Can you write out statements of the principles involved? Although many electric terms, such as voltage, current, and watt, are familiar because they are used in everyday activities, this is not true of magnetic terms. You should not go beyond this point without being sure that you know the precise names and the units of each of the variables in these six relations.

Magnetic Circuits

A magnetic field is usually a means to an end rather than an end in itself. We wish to establish regions of intense magnetic flux density because of the effect of such fields on moving charges. The action of the field can produce beam deflection in an oscilloscope, torque in a motor, or voltage in a generator. By an optimum arrangement of coil and ferromagnetic material the engineer provides the necessary magnetic field at the lowest cost or with a minimum weight.

The Magnetic Circuit Concept

In the electric circuit of Fig. 21.7a, current I is proportional to voltage V and the constant of proportionality is the conductance $G = 1/R$ (Eq. 11-4); the conductance depends on the geometry of the conducting path and a property called conductivity. Equation 21-14 indicates that flux ϕ is proportional to magnetomotive force $\mathfrak{F}$ and the constant of proportionality depends on the geometry of the magnetic path and a property called permeability. The analogy is emphasized if we write the relations for Fig. 21.7 as follows:



(a) (b)

Figure 21.7 Electric and magnetic circuits.

$$\frac{V}{I} = R = \frac{1}{\sigma}\frac{l}{A} \qquad \frac{\mathfrak{F}}{\phi} = \mathcal{R} = \frac{1}{\mu}\frac{l}{A} \tag{21-15}$$

The ratio of mmf $\mathfrak{F}$ to flux ϕ is $\mathcal{R}$, the *reluctance* in ampere-turns/weber. In many practical situations the flux density is uniform and a circuit approach, using a calculated reluctance, is possible. (See Example 3 on the next page.)

It must be emphasized that nothing "flows." in a magnetic circuit. Another difference is that in ferromagnetic materials permeability varies widely with flux density, whereas in most conductors conductivity is independent of current density within the normal operating range. In spite of these differences, the magnetic circuit concept is very useful. The analogous quantities are shown in Table 21-1 on the following page.

Example 3

Use a circuit approach to calculate the current required to establish the flux of 0.2 mWb in the iron core of Example 2.

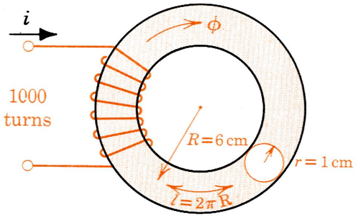


Figure 21.8 A magnetic circuit.

First the reluctance of the path is calculated. By Eq. 21-15,

$$\mathcal{R} = \frac{1}{\mu}\frac{l}{A} = \frac{1}{2000 \times 4\pi \times 10^{-7}}\,\frac{2\pi \times 0.06}{\pi(10^{-2})^2}$$
$$= 4.75 \times 10^5 \text{ A-t/Wb}$$

Then

$$\mathcal{F} = Ni = \phi\mathcal{R} = 2 \times 10^{-4} \times 4.75 \times 10^5 = 95 \text{ A-t}$$

and the current required in a 1000-turn coil is

$$i = \frac{\mathcal{F}}{N} = \frac{95 \text{ A-t}}{1000 \text{ t}} = 0.095 \text{ A} = 95 \text{ mA}$$

Magnetic Circuit Calculations

The circuit approach is particularly useful if the magnetic field is confined to paths of simple geometry and if the flux density is uniform within each component of the path. In the relay of Fig. 21.9a, the coil wound on a *core* 1 establishes a magnetic flux that is largely confined to a path consisting of a fixed iron *yoke* 2, a movable iron *armature* 3, and an *air gap* 4. Half of a four-pole generator is shown in Fig. 21.9b. Flux ϕ_1 follows a path consisting of a pole structure N, an air gap, a yoke, another air gap, and a pole structure S. In the complete generator, there are four similar magnetic circuits.

Reasoning by analogy from electric circuits, we conclude that:

For magnetic circuit elements in series,

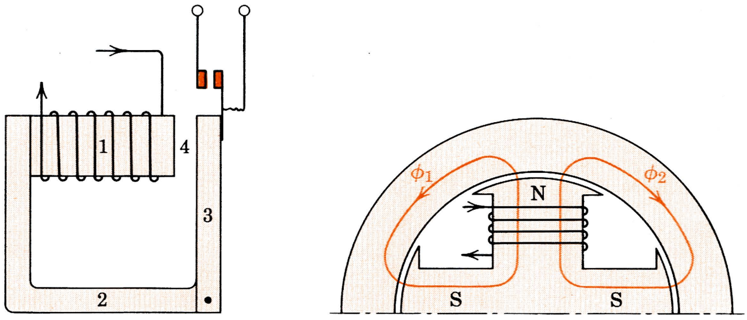
$$\phi_1 = \phi_2 = \cdots = \phi_n \quad \text{and} \quad \mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2 + \cdots \mathcal{F}_n \qquad (21\text{-}16)$$

For magnetic circuit elements in parallel,

$$\phi = \phi_1 + \phi_2 + \cdots + \phi_n \quad \text{and} \quad \mathcal{F}_1 = \mathcal{F}_2 = \cdots = \mathcal{F}_n \qquad (21\text{-}17)$$

Table 21-1 Electric and Magnetic Circuit Analogies

Electric		Magnetic	
Current density	J	Magnetic flux density	B
Current	I	Magnetic flux	ϕ
Electric field intensity	$\mathcal{E}$	Magnetic field intensity	H
Voltage	V	Magnetomotive force	$\mathcal{F}$
Conductivity	σ	Permeability	μ
Resistance	R	Reluctance	$\mathcal{R}$
$\mathcal{E}l = V = IR = \frac{J}{\sigma}l$		$Hl = \mathcal{F} = \phi\mathcal{R} = \frac{B}{\mu}l$	



(a) Relay (b) Four-pole generator

Figure 21.9 Devices incorporating series and parallel magnetic circuits.

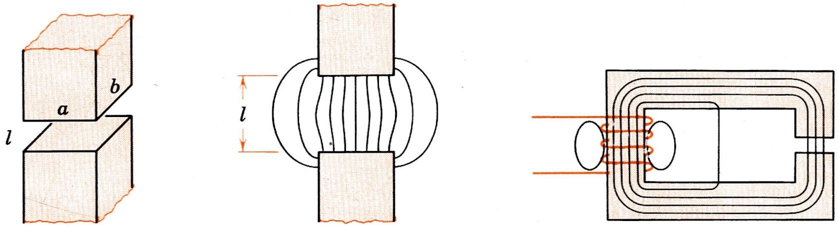
In magnetic circuits the "source" of mmf is usually a coil carrying a current. (It could be a permanent magnet.) The "applied" mmf is just equal to the sum of the mmf "drops" across the elements of a series magnetic circuit. The mmf "drops" across parallel elements are equal; the total flux is the sum of the fluxes in parallel elements.

Series Magnetic Circuits

A common problem is to find the current required to establish a given flux distribution. The procedure is as follows:

1. Analyze the magnetic circuit into a combination of elements in which the flux density is approximately uniform.
2. Determine the flux density $B = \phi/A$ in each element and the corresponding field intensity H.
3. Calculate the mmf drops $\mathfrak{F} = Hl$ in each element.
4. Calculate the total mmf and the required current.

The calculation of the mmf drop across an air gap is straightforward if it is assumed that the flux density is the same as that in the adjoining iron elements. Actually, there is always some *fringing* of magnetic flux (Fig. 21.10a) and the flux density in the gap is lower than in the adjacent iron. One approximate rule for accounting for fringing in short gaps is to increase each dimension (a and



(a) (b)

Figure 21.10 Airgap fringing and leakage flux.

b) by the length of the gap. However, neglecting fringing gives a conservatively high value for the mmf drop and simplifies the calculation. Some flux produced by a coil may return through a short air path (Fig. 21.10b) and never reach the air gap. This *leakage flux* is usually small if the permeability of the iron is high. As a result of flux leakage, a higher mmf is required to establish a given flux in the air gap.

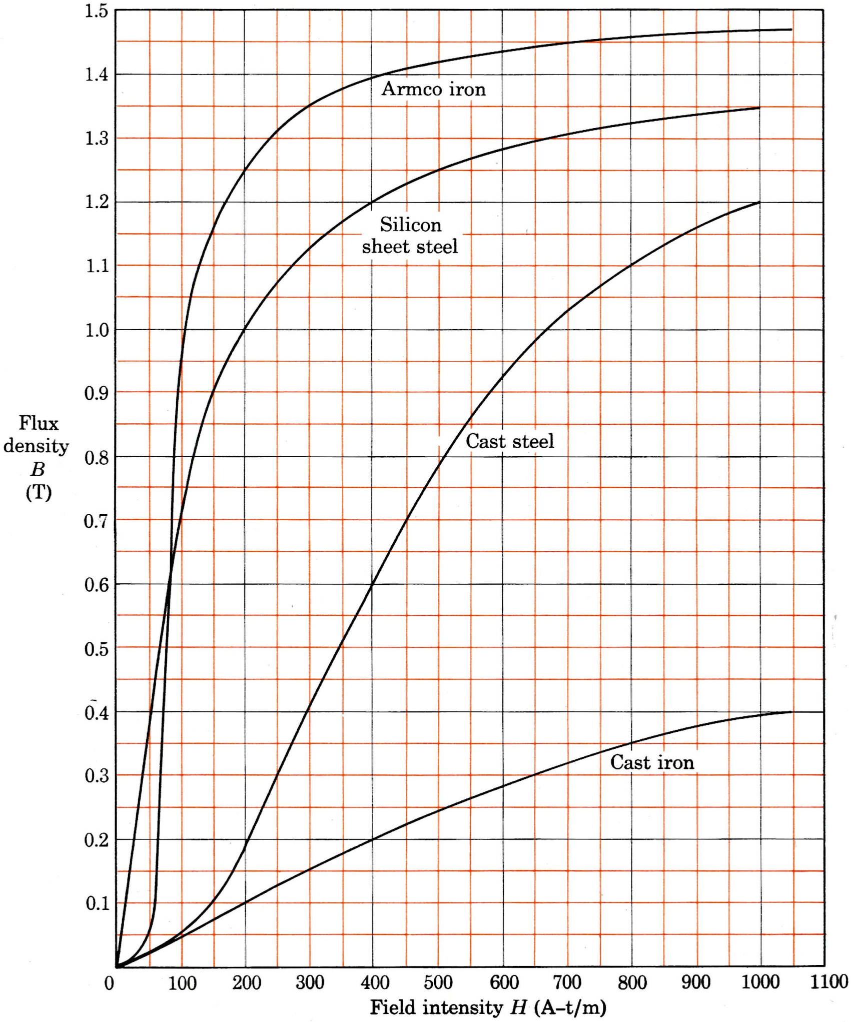


Figure 21.11 Typical magnetization curves for common magnetic materials.

The mmf drop across an iron section can be calculated by assuming a value of relative permeability or by working from average data in the form of *magnetization curves*. Curves similar to those in Fig. 21.11 are available from manufacturers of magnetic materials. It must be emphasized that these are typical curves and results based on such data are not precise. In fact, variations in values of magnetic properties of 3 to 5% are to be expected and there is no point in attempting more precise calculations.

Example 4

Given the magnetic circuit shown in Fig. 21.12 with 500 turns wound on each leg, find the current required to establish a flux of 4 mWb across the 0.1-cm air gaps.

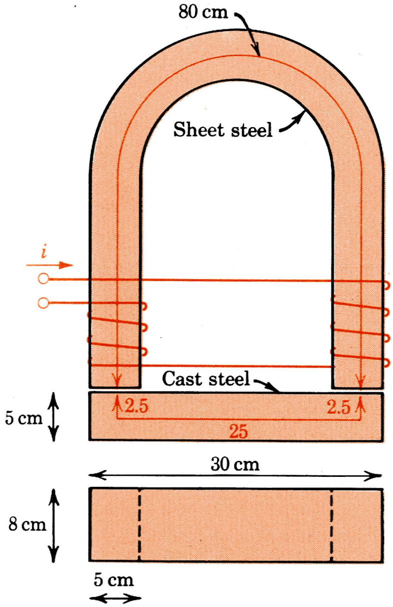


Figure 21.12 A series magnetic circuit.

Following the procedure outlined, we consider the magnetic circuit to consist of two iron elements and two air gaps in series. We assume that fringing is negligible and that the effective length of the cast steel element is 25 + 2.5 + 2.5 = 30 cm; the increased flux density at the inner corner is partially offset by the decreased flux density at the outer corner.

Sheet steel: $$B = \frac{\phi}{A} = \frac{4 \times 10^{-3}\ \text{Wb}}{0.05 \times 0.08\ \text{m}^2} = 1\ \text{T}$$

From Fig. 21.11, $H = 200$ A-t/m and

$$\mathfrak{F} = Hl = 200 \times 0.8 \cong 160\ \text{A-t}$$

Cast steel: $$B = \frac{\phi}{A} = \frac{4 \times 10^{-3}\ \text{Wb}}{0.05 \times 0.08\ \text{m}^2} = 1\ \text{T}$$

From Fig. 21.11, $H = 670$ A-t/m and

$$\mathfrak{F} = Hl = 670 \times 0.3 \cong 200\ \text{A-t}$$

Air gaps: $$\mathfrak{F} = 2Hl = 2\frac{B}{\mu}l = \frac{2\phi l}{A\mu_0}$$

$$= \frac{2(0.004)(0.001)}{0.004(4\pi \times 10^{-7})} = 1590\ \text{A-t}$$

By Eq. 21-16, the total mmf required is

$$160 + 200 + 1590 = 1950\ \text{A-t}$$

Since the mmfs of the two coils are in the same direction, the effective number of turns is 1000 and the current required is

$$i = \frac{\mathfrak{F}}{N} = \frac{1950}{1000} = 1.95\ \text{A}$$

Note that in Example 4 the high permeability of the iron results in a small mmf drop even though the path length is great compared to the length of the air gap. Since the mmf drop across the air gap is the predominant factor, as a first approximation the other mmf drops can be neglected. As a second approximation, the iron elements can be taken into account using average values of permeability.

Example 5

Repeat Example 4, using the first approximation described in the preceding paragraph.

Assuming all the reluctance is in the air gap,

$$i = \frac{\mathcal{F}}{N} = \frac{1590}{1000} \cong 1.6\ \text{A}$$

From Fig. 21.11, the average permeabilities for $B \leq 1$ T are

Sheet steel:

$$\mu_r = \frac{B}{\mu_o H} \cong \frac{1}{4\pi \times 10^{-7} \times 200} \cong 4000$$

Cast steel:

$$\mu_r = \frac{B}{\mu_o H} \cong \frac{1}{4\pi \times 10^{-7} \times 700} \cong 1000$$

Repeat Example 4, using the second approximation.

For a series magnetic circuit with constant μ,

$$\mathcal{R} = \mathcal{R}_a + \mathcal{R}_{ss} + \mathcal{R}_{cs} = \frac{1}{\mu_o A}\left(\frac{l_a}{\mu_a} + \frac{l_{ss}}{\mu_{ss}} + \frac{l_{cs}}{\mu_{cs}}\right)$$

$$= \frac{1}{4\pi \times 10^{-7}(0.004)}\left(\frac{0.002}{1} + \frac{0.8}{4000} + \frac{0.3}{1000}\right)$$

$$\cong 5 \times 10^5\ \text{A-t/Wb}$$

and

$$i = \frac{\mathcal{F}}{N} = \frac{\phi\mathcal{R}}{N} \cong \frac{4 \times 10^{-3} \times 5 \times 10^5}{1000} \cong 2\ \text{A}$$

We conclude that the method to be used in a given problem depends on the data available and the precision required.

Series Circuit with Given Mmf

In Example 4 the problem is to find the mmf required for a given flux distribution. Knowing the flux density in each element of the circuit, we can determine the corresponding field intensity and mmf. Finding the precise flux distribution resulting from a given mmf is a more difficult problem because of the nonlinearity of the magnetization curves. (The approximate methods used in Example 5 assume linearity and work equally well for both classes of problems.)

The methods of analysis outlined in Chapter 2 are applicable to nonlinear magnetic circuits. The method of piecewise linearization (see Fig. 2.26a) is essentially the approach used in Example 5, where we assumed average values of relative permeability. If there is a single nonlinear element, a load line can be constructed as in Fig. 2.27. (If a magnetic circuit consists of an iron element, an air gap, and a multiturn coil, what are the analogs of V_T, R_T, and R_N?) An effective approach is first to obtain a rough approximation by assuming all the reluctance is in the air gap and then to proceed by "educated" trial and error.

Example 6

The core of cast steel in Fig. 21.13 has a cross-section of 10 cm^2 and an average length of 35 cm with a 1-mm air gap. It is wound with 200 turns of wire carrying a current of 3 A. Determine the total flux across the gap.

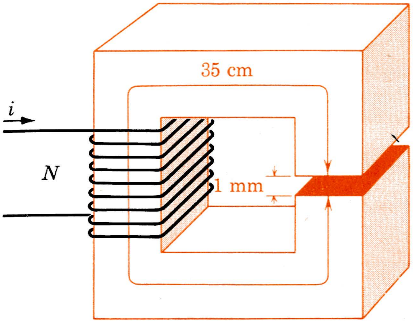


Figure 21.13 Magnetic flux calculation.

As a first approximation, assume that all the mmf is across the air gap. Then

$$B = \mu H = \frac{\mu_o \mathfrak{F}}{l} = \frac{\mu_o N i}{l} = \frac{4\pi \times 10^{-7} \times 200 \times 3}{10^{-3}}$$

$$= 0.75\ \mathrm{T}$$

The actual value of B is less than this value because of the mmf drop in the cast steel. Assuming $B = 0.6$ T,

$$\mathfrak{F}_a = H_a l_a = \frac{B l_a}{\mu_o} = \frac{0.6 \times 10^{-3}}{4\pi \times 10^{-7}} = 478\ \text{A-t}$$

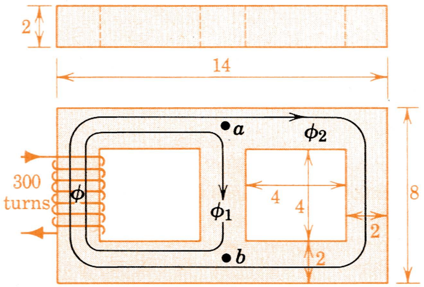
From Fig. 21.11, $H_{cs} \cong 400$ A-t/m and
$\mathfrak{F}_{cs} = H_{cs} l_{cs} = 400 \times 0.35 = 140$ A-t.
The total mmf required is $478 + 140 = 618$ A-t. This is within 3% of the $3(200) = 600$ A-t available and, therefore, $B = 0.6$ T is an acceptable value. The total flux is $\phi = BA = 0.6 \times 10^{-3} = 0.6$ mWb. (Note: If greater precision were required, a second guess would be that $B_2/B_1 = \mathfrak{F}_2/\mathfrak{F}_1 = 600/618$.)

Parallel Magnetic Circuits

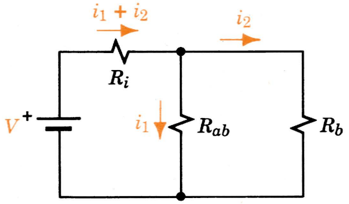
The analysis of magnetic circuits containing parallel elements is based on analogy with the corresponding electric circuits. The same mmf (voltage) exists across elements in parallel and the total flux (current) is the sum of the fluxes (currents) in the parallel elements. The procedure is illustrated in Example 7.

Example 7

A flux of 3.6 × 10^{-4} Wb is to be established in the center leg of the sheet steel core shown in Fig. 21.14 (all dimensions in cm). Find the necessary current in the 300-turn coil.



(a) Magnetic circuit



(b) Electrical analogy

Figure 21.14 Parallel circuit calculations.

Possible flux paths have been sketched in to indicate the general distribution of flux. Assuming that flux density between points *a* and *b* is approximately uniform, the flux density in the center leg is

$$B_1 = \frac{\phi_1}{A_1} = \frac{3.6 \times 10^{-4}}{0.02 \times 0.02} = 0.9\ \mathrm{T}$$

From Fig. 21.11, for $B_1 = 0.9$ T, $H_1 = 150$ A-t/m and

$$\mathfrak{F}_{ab} = H_1 l_1 \cong 150 \times 0.06 = 9\ \text{A-t.}$$

Then

$$H_2 = \frac{\mathfrak{F}_{ab}}{l_2} \cong \frac{9}{(6 + 6 + 6)10^{-2}} = \frac{9}{0.18} = 50\ \text{A-t/m}$$

From Fig. 21.11, for $H_2 = 50$ A-t/m, $B_2 \cong 0.35$ T and

$$\phi_2 = B_2 A_2 = 0.35 \times 4 \times 10^{-4} = 1.4 \times 10^{-4}\ \mathrm{Wb}$$

To the left of *ab*, the total flux is

$$\phi = \phi_1 + \phi_2 = (3.6 + 1.4)10^{-4} = 5 \times 10^{-4}\ \mathrm{Wb}$$

and

$$B = \frac{\phi}{A} = \frac{5 \times 10^{-4}}{4 \times 10^{-4}} = 1.25\ \mathrm{T}$$

From Fig. 21.11, for $B = 1.25$ T, $H = 500$ A-t/m and

$$\mathfrak{F} = Hl = 500 \times 0.18 = 90\ \text{A-t.}$$

The required current in the 300-turn coil is

$$i = \frac{\mathfrak{F} + \mathfrak{F}_{ab}}{N} = \frac{90 + 9}{300} = 0.33\ \mathrm{A}$$

SUMMARY

- Magnetic flux density **B** is defined in terms of the force exerted on a moving charge by the equation $\mathbf{f} = q\mathbf{u} \times \mathbf{B}$.
  The contribution of an element *ds* carrying a current *i* is

$$dB = \frac{\mu i\, ds \cos\alpha}{4\pi r^2}\ \text{teslas}$$

- Magnetic flux $\phi = \int \mathbf{B} \cdot d\mathbf{A}$ webers = summation of flux density.
  Magnetic flux lines are continuous or $\oint \mathbf{B} \cdot d\mathbf{A} = 0$.

- The permeability of free space is $\mu_o = 4\pi \times 10^{-7}$ henrys/meter.
  The relative permeability of a material is $\mu_r = \mu/\mu_o$.

- Magnetic field intensity is defined by $\mathbf{H} = \dfrac{\mathbf{B}}{\mu}$ amperes/meter (or A-t/m).

  The line integral of magnetic field intensity along any closed path is just equal to the current linked or $\oint \mathbf{H} \cdot d\mathbf{l} = i$ amperes (or A-t).

- Ferromagnetism is due to the magnetic effects of unbalanced electron spins and the behavior of domains under the action of external magnetic fields. The resulting magnetic polarization M is defined by $B = \mu_o(H + M) = \mu_o\mu_r H$. Magnetization curves provide B–H data for magnetic materials.

- Analogies between magnetic and electric quantities provide the basis for a circuit approach to magnetic field problems.

  Series: $\phi_1 = \phi_2 = \cdots = \phi_n$ and $\mathfrak{F} = Ni = \Sigma Hl$
  Parallel: $\mathfrak{F}_1 = \mathfrak{F}_2 = \cdots = \mathfrak{F}_n$ and $\phi = \Sigma BA$

- To find the current required to establish a given magnetic flux:

  1. Subdivide the circuit into uniform elements.
  2. Determine B in each element and the corresponding H.
  3. Calculate the mmf drops $\mathfrak{F} = Hl$ in each element.
  4. Calculate the total mmf and the required current.

- To find magnetic flux for a given current, use graphical analysis, piecewise linearization, or trial and error.
  Magnetic calculations are never precise; in a practical device with an air gap the mmf drop in the iron may be negligible.

REVIEW QUESTIONS

1. What is a magnetic field? What are its properties?
2. How could you determine experimentally if a magnetic field is present?
3. What are lines of magnetic force? How are they useful?
4. What is the flux density at the center of a long, straight, copper conductor carrying current i? A similar iron conductor?
5. Why is iron magnetic and aluminum is not?
6. Why does relative permeability vary with flux density?
7. Sketch the flux distribution around a straight conductor.
8. Explain Equation 21.11 for a permanent magnet.
9. Given an $i\,ds$ element, what factors determine dB at point P?
10. Sketch analogous magnetic and electric circuits and list four analogous terms. How are these circuits basically different?
11. List five magnetic field quantities with their symbols and units.
12. In magnetic circuits, what corresponds to Kirchhoff's laws?
13. Explain the statement: "Applied mmf equals the sum of the mmf drops."
14. Under what circumstances can all the mmf be assumed across the air gap?
15. Given the current in a coil wound on a magnetic core with an air gap, outline the procedure for finding total flux.

EXERCISES

1. Sketch a conducting element of length l carrying a current i consisting of carriers of charge q and velocity u. Taking into account area A, carrier density n, and total charge Q, demonstrate that $Qu = il$.
2. A current i in the short conducting element shown in Fig. 21.15 produces a flux density B_1 at point 1. Determine the magnitude and the direction of the flux density vectors at points 2, 3, and 4.

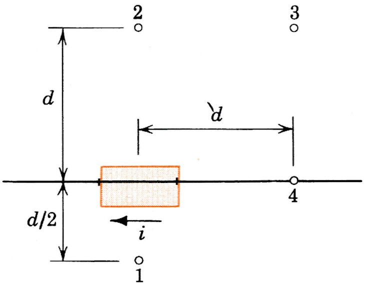


Figure 21.15

3. A circular coil of radius R carrying a current I lies in the x-y plane with its center at the origin. Derive an expression for the flux density B as a function of distance along the z axis.
4. A 10-turn compact circular coil 10 cm in diameter is to provide a flux density of 0.5 mT at its center. Specify the required current.
5. Demonstrate that the units of permeability are henrys/meter.
6. In a certain material the flux density is 1.0 T. Calculate the magnetic field intensity and the magnetic polarization if the relative permeability of the material is: (a) 1.00002 and (b) 2500.
7. A straight conductor carrying a current I passes through the center of a square 1 cm on a side, in a plane parallel to two sides of the square and at an angle of 30° with respect to the normal to the square. Determine the total mmf around the square.
8. One commonly used memory element for digital computers is a small toroidal core in which the direction of magnetization distinguishes between 0 and 1. if the core (Fig. 21.6) is of *ferrite*, for which $\mu_r = 1200$, and $R = 0.2$ cm and $r = 0.05$ cm, estimate the magnetomotive force required to establish a flux density of 1 millitesla.
9. A straight conductor 1 m long and 2 mm in diameter carries a current of 10 A. Determine and plot the magnitude of the magnetic flux density in the air around the conductor.
10. The long solenoid coil of Fig. 21.16 is one way of getting a uniform flux density over a region. Recognizing this solenoid as a special form of the toroid of Fig. 21.6, the analysis is similar.
    (a) Evaluate $\int H dl$ along paths a and b.
    (b) Assuming an infinitely long solenoid, evaluate $\int H dl$ along path dL and express B in that region.
    (c) For an "air-core" solenoid of length $L = 20$ cm, $r = 1$ cm, and $N = 100$ turns, predict the flux density and total flux across the median section for a current of 2 A.

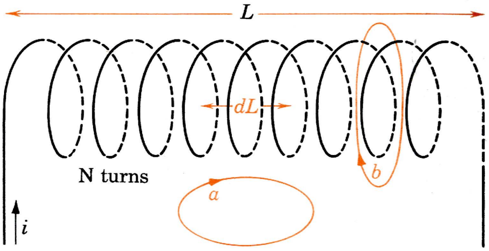


Figure 21.16

11. Iron with a relative permeability of 1500 is formed into a toroidal core with $R = 12$ cm and $r = 1$ cm. Design a coil to provide a flux density of 0.8 T in the core.
12. Refer to the six relations on page 556.
    (a) Which are *definitions*, and what is defined in each?
    (b) Write out in words the principles expressed in the *experimental* relations.

13. (a) Derive an expression for the equivalent reluctance $\mathcal{R}_{EQ}$ of a magnetic circuit consisting of reluctances $\mathcal{R}_a$, $\mathcal{R}_b$, and $\mathcal{R}_c$ in series.

(b) Repeat part (a) for reluctances in parallel.

14. Calculate the average value of relative permeability for flux densities up to $B = 0.4$ T for: (a) Cast iron, (b) Cast steel, (c) Silicon sheet steel, and (d) Armco iron.

15. (a) Calculate about five values and plot a graph of relative permeability versus flux density for cast steel.

(b) Repeat for Armco iron.

16. For experiments in space propulsion, engineers developed a 6-inch diameter solenoid providing a flux density of 10.7 T across a 1-inch bore by using a superconducting magnet immersed in helium. Current is initiated by a 6-V battery and then continues without degradation. Estimate the current required in a 1000-turn coil to establish this flux density across 10 cm of air gap.

17. (a) A symmetric core of silicon sheet steel is similar to that in Fig. 21.17 but with no air gap. If $a = c = 5$ cm and $b = d = 25$ cm, find the total magnetic flux produced by a current of 2.5 A in the 200-turn coil.

(b) Repeat for a core of cast steel.

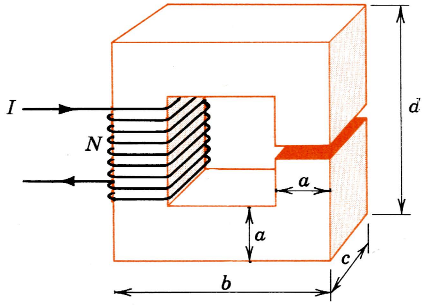


Figure 21.17

18. A symmetric core similar to that in Fig. 21.17, but with no air gap, is to carry a total flux of 1.6 mWb. Dimensions are $a = c = 4$ cm, and $b = d = 20$ cm.

(a) Estimate the current required in a 100-turn coil if the core is Armco iron.

(b) Repeat for a cast steel core.

(c) Repeat for a cast iron core.

19. The core of Fig. 21.17 is Armco iron with $a = c = 10$ cm, and $b = d = 40$ cm; the current in the 300-turn coil is 10 A.

(a) Estimate the allowable air gap length for a magnetic flux of 12 mWb.

(b) Estimate the error in percent that would result from neglecting the mmf drop in the iron.

20. Armco iron is used in the core of Fig. 21.17 with $a = c = 10$ cm and $b = d = 50$ cm. You are to estimate the total flux across the 3-mm air gap when the current is 1.5 A in the 2000-turn coil.

(a) What assumption could be made to obtain an approximate value of B? Make such an assumption and calculate B.

(b) Is the actual value larger or smaller than this B?

(c) Assume a value of B and calculate the corresponding I.

(d) Compare this value of I with the given current. If they agree within 5%, the problem is solved; if not, repeat parts (c) and (d).

21. Silicon sheet steel is used in the core of Fig. 21.17 with $a = 4$ cm, $b = 20$ cm, $c = 5$ cm, and $d = 25$ cm. Predict the flux density in the 2-mm gap with a current of 5.5 A in the 400-turn coil.

22. Silicon sheet steel is used in the core of Fig. 21.18 with $a = 20$ cm, $b = 40$ cm, $c = 120$ cm, and $d = 60$ cm. $N_1 = N_2 = 500$ turns. Estimate the value of $I_1 = I_2$ to produce a total magnetic flux of 100 mWb across the 5-mm air gap. (Hint: Can symmetry be used to simplify this problem?)

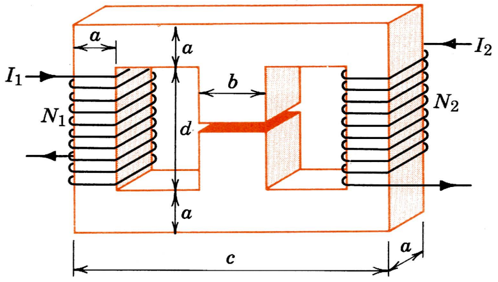


Figure 21.18

23. A single 500-turn coil on the center leg of the core of Exercise 22 replaces the two coils on the outer legs. Without detailed calculation, predict the required current I in terms of I_1. Which is the more economic design?

24. Cast steel is used in the core of Fig. 21.18 with $a = 5$ cm, $b = 8$ cm, $c = 40$ cm, and $d = 20$ cm. $N_1 = N_2 = 450$ turns; $I_1 = 5$ A and $I_2 = -5$ A.
   (a) Draw the analogous electric circuit.
   (b) Sketch the flux distribution showing lines close together where flux density is high, etc.
   (c) Estimate the flux density in the 2-mm air gap.

PROBLEMS

1. Starting from Ampere's law (Eq. 21-5), derive an expression for the flux density at the center of a square coil L meters on a side carrying a current I. Repeat for a compact coil of N turns.

2. One way to obtain a nearly uniform magnetic field in air is to use two compact circular N-turn coils of radius R spaced a distance R apart along their common axis.
   (a) Using the result of Exercise 3, derive an expression for the flux density B at the geometric center of the combination.
   (b) For 100-turn coils of 10-cm radius carrying 5 A, calculate B at the geometric center of the combination and at the midpoint of one coil and draw a conclusion.

3. A core of silicon sheet steel has the general shape of Fig. 21.17. Stating all assumptions, determine a ratio of effective length of iron l_i to effective length of air l_a to define the conditions under which the mmf drop in the iron can be neglected without introducing an error greater than 5% in the flux calculations.

4. The core of Fig. 21.17 is silicon sheet steel with $a = c = 5$ cm and $b = d = 30$ cm. The 400-turn coil carries a current of 15 A. Estimate the total flux across the 5-mm air gap by using a "load line" drawn on a graph of ϕ vs. $\mathfrak{F}$ for the core.

5. Design the winding for a high permeability silicon steel core in the shape of Fig. 21.17 to provide a flux density of 1 T across the 5 mm air gap ($a = 5$ cm and $c = 10$ cm). Assume that copper windings occupy half the area of the "window" in the core, the remainder being taken up with insulation and air space, and that the allowable current density in copper is 200 A/cm^2.
   (a) Determine the area of the window and specify the dimensions of the core.
   (b) Estimate the mean turn length and specify the wire resistance per meter for operation at 120 V dc.
   (c) Select an appropriate wire size from a table of copper wire gages.

6. A core similar to that in Fig. 21.17 but without an air gap is placed around the neck of a TV picture tube to provide magnetic deflection. A second coil identical to that shown is wound on the right-hand leg; it carries the same current I in such a direction as to produce a *counterclockwise* mmf. Sketch the resulting magnetic field and describe qualitatively the deflection obtained.

Chapter 22
Transformers

AC Excitation of Magnetic Circuits
Transformer Operation
Linear Circuit Models
Transformer Performance

In 1831 Michael Faraday applied a voltage to one coil and observed a voltage across a second coil wound on the same iron core. Fifty-three years later at an exhibition in Turin, Italy, the English engineers Gaulard and Gibbs demonstrated an "electric system supplied by inductors," using "secondary generators" with open iron cores.† Three young Hungarian engineers, Déri, Bláthy, and Zipernowsky, who visited the Turin exhibition, recognized the disadvantages of the open core and started work on an improved version. Five weeks later they shipped their "Transformer No. 1," a 1400-W model with a closed core of iron wires. Seventy-five of their transformers were used to supply the 1067 Edison lamps that illuminated the 1885 Budapest exhibition. In one observer's words: "A new system of distribution has been inaugurated which . . . bids fair to mark an epoch . . . . It becomes possible to conduct the current from one central station to many consumers, even over very great distances." That prediction has come true and the transformer is a key energy-conversion device in modern power-distribution systems.

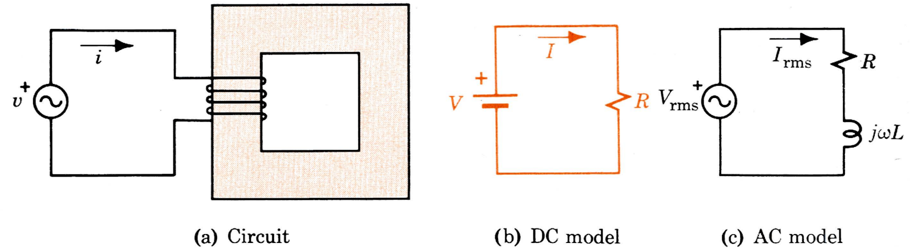
The transformer transfers electrical energy from one circuit to another by means of a magnetic field that links both circuits. There are three reasons for investigating its behavior at this point in our studies: The transformer illustrates the use of magnetic circuits in energy conversion, it provides a good example of voltages induced by a changing magnetic field, and transformer action is the basis of operation of induction motors and other important electromechanical devices. First we consider magnetic circuits with sinusoidal flux variation, then we investigate the voltage and current relations in coils linking the same magnetic circuit. Next we treat the transformer as a two-port and derive a circuit model. Finally we use the model to predict transformer performance.

---

† Halacsy-von Fuchs: "Transformer Invented 75 Years Ago," *Electrical Engineering*, June 1961, p. 404.

AC Excitation of Magnetic Circuits

When a steady or dc voltage exists across a coil linking a magnetic core, the current that flows is limited by the *resistance* of the winding. The voltage RI developed across the resistance is just equal to the applied voltage V and the device of Fig. 22.1a can be represented by the circuit model of Fig. 22.1b. The magnetic flux ϕ established in the core is determined by the magnetizing force H in ampere-turns per meter, the properties of the core as defined by a B–H curve, and the cross-sectional area A of the core.



(a) Circuit (b) DC model (c) AC model

Figure 22.1 Magnetic circuit excited by dc and ac voltages.

In contrast, when an ac voltage exists across the same coil, the current is limited by the *impedance* of the winding. The voltage drop due to resistance is usually small, and the voltage drop due to inductive reactance is approximately equal to the applied voltage. The magnetic flux in the core is just that required to produce an induced voltage approximately equal to the applied voltage. In the circuit model of Fig. 22.1c, this effect is represented by an inductance L. Originally we described inductance as a measure of energy stored in a magnetic field, but we defined it as a circuit quantity in terms of its v-i characteristic. Now we are in a position to consider inductance in terms of the magnetic field that it represents.

Induced Voltage and Inductance

On the basis of his experiments, Faraday correctly concluded that the voltage induced in a multiturn coil linking a changing magnetic field is proportional to the number of turns N and to the time rate of change of flux ϕ or

$$v = N\frac{d\phi}{dt} = \frac{d\lambda}{dt} \tag{22-1}$$

where λ (lambda) is the number of *flux linkages* in weber-turns. If, in a 10-turn coil, 8 turns link a flux of 1 Wb and 2 turns link a flux of only 0.9 Wb, the total flux linkage is 8 × 1 plus 2 × 0.9 or 9.8 Wb-t. (Here again the number of turns is a dimensionless quantity, but we carry the term in the unit as a reminder of its physical significance.)

The polarity of the induced voltage can be determined by *Lenz's law:*

> *The induced voltage is always in such a direction as to tend to oppose the change in flux linkage that produces it.*

Consider, for example, the case of a coil linked by a *decreasing* flux. If a closed path is provided, the current caused by the induced voltage will be in such a direction as to produce *additional* flux.

In general, the relation between flux and current is nonlinear; however, in many practical situations linearity can be assumed with negligible error. If the flux in a magnetic circuit varies sinusoidally so that

$$\phi = \Phi_m \sin \omega t$$

the induced voltage in an N-turn coil is

$$v = N\frac{d\phi}{dt} = N\omega\Phi_m \cos \omega t \tag{22-2}$$

The rms value of this sinusoidal voltage is

$$V = \frac{V_m}{\sqrt{2}} = \frac{N\omega\Phi_m}{\sqrt{2}} = \frac{2\pi}{\sqrt{2}}Nf\Phi_m = 4.44Nf\Phi_m \tag{22-3}$$

For ac excitation, the resulting flux Φ_m depends on the frequency as well as the magnitude of the applied voltage. If the resistance drop is neglected, the current which flows is just that required to establish the flux specified by Eq. 22-3.

By the definition of inductance, the voltage developed is

$$v = L\frac{di}{dt} \tag{22-4}$$

Combining this definition with Faraday's law,

$$L\frac{di}{dt} = N\frac{d\phi}{dt} = \frac{d\lambda}{dt} \qquad \text{or} \qquad L = N\frac{d\phi}{di} = \frac{d\lambda}{di} \tag{22-5}$$

For a linear magnetic circuit, ϕ is proportional to i or

$$L = \frac{N\phi}{i} = \frac{\lambda}{i} \tag{22-6}$$

and another interpretation of this circuit parameter is:

> *Inductance is the number of flux linkages per unit of current.*

Example 1

A current of 2 A in a coil produces a magnetic field described by the flux pattern shown in Fig. 22.2 where each line represents 1 mWb of flux. Determine the flux linkage created and the inductance of the coil.

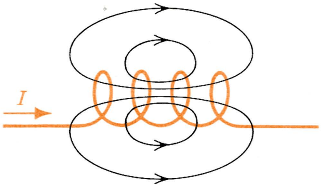


Figure 22.2 Flux linkages.

Since 2 coil turns link 4 mWb and the other 2 turns link only 2 mWb, the total flux linkage is

$$\lambda = 2 \times 4 \times 10^{-3} + 2 \times 2 \times 10^{-3} = 12 \text{ mWb-t}$$

By Eq. 22.6 the inductance is

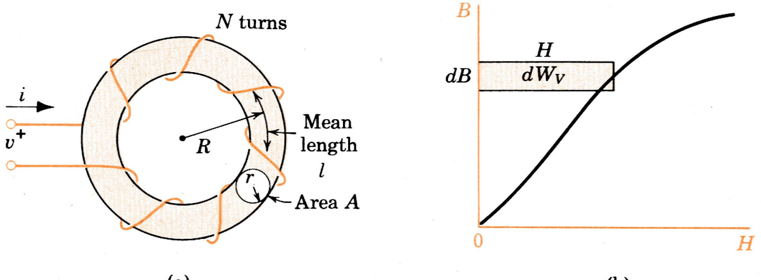
$$L = \frac{\lambda}{i} = \frac{12 \times 10^{-3}}{2} = 6 \times 10^{-3} \text{ H} = 6 \text{ mH}$$

In Example 1 we see the important factors in designing an inductance. If many turns are wound on a high-permeability core, the flux linkage per ampere is high and the inductance is correspondingly great.

Energy Storage in a Magnetic Field

Another interpretation of inductance is as a measure of the ability of a circuit component to store energy in a magnetic field (Eq. 1-15). This is an important concept because the magnetic field is frequently a coupling medium in transforming energy from one form to another. How can the stored energy be calculated?

In the toroidal core of Fig. 22.3a the mmf is $\mathfrak{F} = Ni = Hl$. Assuming that the core is "thin" ($r \ll R$), flux density B and magnetizing force H are uniform across the area A and the total flux is $\Phi = BA$. Starting with an unmagnetized core, we can store energy by building up a current and creating a magnetic



(a) (b)

Figure 22.3 Energy storage in a magnetic field.

field. The magnetic field energy stored comes from the electric circuit, and the electrical input is

$$W = \int_0^t vi\,dt = \int_0^t N\frac{d\phi}{dt}\cdot i\,dt = \int_0^\phi Ni\,d\phi \tag{22-7}$$

But $Ni = \mathfrak{F} = Hl$ and $d\phi = A\,dB$, therefore

$$Ni\,d\phi = (Hl)(A\,dB) = (lA)H\,dB$$

where lA is the volume of the magnetic core. The energy stored per unit volume, i.e., the *energy density*, is

$$W_V = \frac{W}{lA} = \int_0^B H\,dB \tag{22-8}$$

which can be interpreted as the area between the magnetization curve and the B axis (Fig. 22.3b). Assuming linear magnetic characteristics, i.e., constant permeability, $B = \mu H$ and Eq. 22-8 can be integrated to yield

$$W_V = \int_0^B H\,dB = \int_0^B \frac{B}{\mu}dB = \frac{1}{2}\frac{B^2}{\mu} \tag{22-9}$$

or

$$W_V = \int_0^B H\,dB = \int_0^H \mu H\,dH = \frac{1}{2}\mu H^2 \tag{22-10}$$

Either of these expressions can be used to calculate the energy density in any part of a magnetic field.

Example 2

An N-turn coil (similar to Fig. 22.3a) is wound on a thin toroidal core of relative permeability μ_r, cross-section A, and length l_i with an air gap of length l_a. For a flux density B in the iron, compare the energy densities and the total energies in the iron and in the air.

If the relative permeability is 2000 and if the length of the air gap is 1% of the length of the iron, estimate the fraction of the total energy that is stored in the air.

Assuming uniform flux distribution and neglecting fringing and leakage flux, the flux density in the air is the same as in the iron. By Eq. 22-9,

$$\frac{W_{Va}}{W_{Vi}} = \frac{B^2/\mu_o}{B^2/\mu_o\mu_r} = \mu_r$$

or the energy density in the air in a series magnetic circuit is μ_r times as great as that in the iron.

The ratio of the total energies is

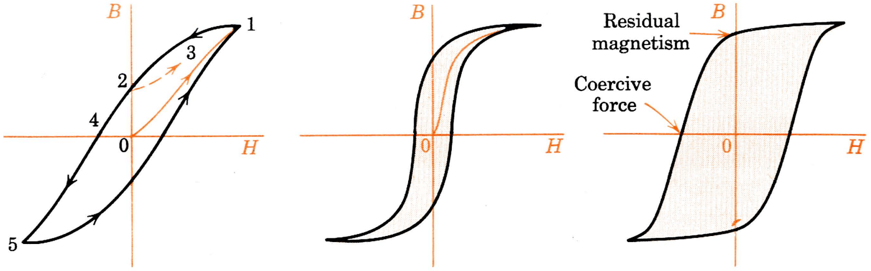
$$\frac{W_a}{W_i} = \mu_r\frac{l_aA_a}{l_iA_i} = \mu_r\frac{l_a}{l_i}$$

$$\frac{W_a}{W_i} = 2000 \times 0.01 = 20$$

Therefore, more than 95% of the energy stored is in the air. In many practical problems all the energy stored can be assumed to be in the air gap.

Hysteresis

In magnetizing a ferromagnetic material by reorientation of the domains, most of the effects are irreversible (inelastic). When the external field is removed, the magnetic material does not return to its original state. If an iron specimen is saturated, point 1 in Fig. 22.4a, and then the field is removed ($H = 0$), the magnetic condition follows the line from 1 to 2. The ordinate at point 2 is called the *residual magnetism.* If a positive H is again applied, the condition follows the path 2 to 3. A negative magnetizing force (the *coercive force*) is required to bring the flux density to zero as at point 4. A large negative H produces saturation in the opposite direction (point 5). Reversing the magnetizing force causes the magnetic condition to follow the path 5 to 1. If the magnetizing force is due to an alternating current, the *hysteresis loop* is traced out once each cycle.



(a) General (b) Silicon steel (c) Alnico

Figure 22.4 Hysteresis loops.

Just as a metal strip gets warm when it is repeatedly flexed, magnetic material gets warm when it is cyclically magnetized. In both cases, the energy that appears as an increase in temperature is due to inelastic action. The energy input

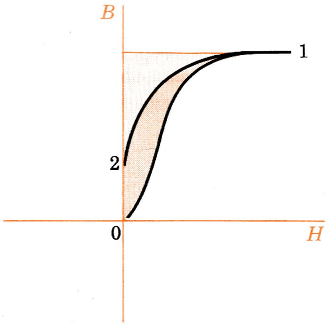


Figure 22.5 Hysteresis loss.

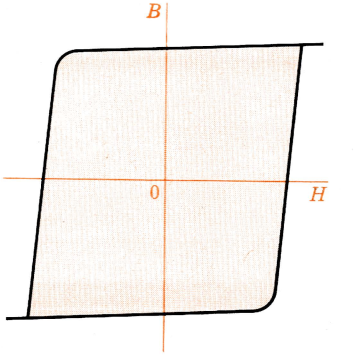


Figure 22.6 Square hysteresis loop.

to an initially unmagnetized sample (of unit volume) is the area between the curve 0–1 and the B axis; the energy returned is the area between the curve 1–2 and the B axis (Fig. 22.5). The difference in these areas is the energy converted to heat in an irreversible process. Following this line of reasoning for a complete cycle, we see that the area of the hysteresis loop is just equal to the energy lost per cycle. For silicon steel the loop is thin (Fig. 22.4b) and the *hysteresis loss* is small; for a "permanent magnet" the coercive force is large and the loop is fat (Fig. 22.4c). The area of the loop increases nonlinearly with maximum flux density. An empirical formula developed by Steinmetz for commercial magnetic steels gives the hysteresis power loss in watts as

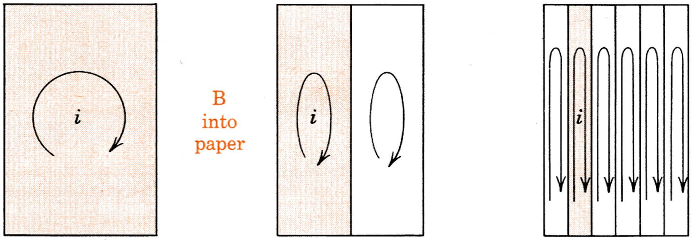
$$P_h = K_h f B_m{}^n \tag{22-11}$$

where the constant K_h and exponent n vary with the core material; n is often assumed to be 1.6.

In some magnetic materials the hysteresis loop is nearly rectangular (Fig. 22.6). In such *square-loop* materials the slope of the sides of the B–H curve is large and a small change in H can *switch* the core from nearly saturated in one direction to nearly saturated in the other. A tiny core and coil can serve as a magnetic flip-flop in switching circuits or as a storage element in a computer.

Eddy Currents

In early electric generators nearly three-fourths of the mechanical input appeared as heat in the magnetic circuit. A small part of this energy loss was due to hysteresis, but the major part was due to the fact that a changing magnetic flux induces voltages in the core material itself. In a conducting iron core, induced voltages cause localized *eddy currents* and the resulting $i^2R = v^2/R$ power appears as heat. As shown in Fig. 22.7a, a changing flux (directed into the paper) induces a net current within the core material. The power loss can be reduced by decreasing v and increasing R. If, instead of a solid iron core, thin *laminations* are used (Fig. 22.7b), the effective induced voltage is decreased and the resistance of the effective path is increased. (What effect does laminating the core have on eddy-current path length? On cross-sectional area?) The laminations, perhaps



(a) (b) (c)

Figure 22.7 Eddy currents and the effect of core lamination.

0.02 in. thick, are electrically insulated from each other by a thin varnish or just the scale produced in heating and rolling.

For a given core, the eddy-current power loss is given by†

$$P_e = K_e f^2 B_m{}^2 \tag{22-12}$$

Since induced voltage is proportional to fB_m (Eq. 22-2) and the loss varies as the square of the voltage, we expect the power loss to vary as $f^2 B_m{}^2$. The constant K_e depends on the resistivity of the core material and the thickness of the laminations. How would you expect K_e to vary with lamination thickness? (See Problem 2.)

Example 3

An iron-core inductor (Fig. 22.8) is designed to operate at 120 V at 60 Hz. Estimate the effect on hysteresis and eddy-current losses of operating at 150 V at 50 Hz.

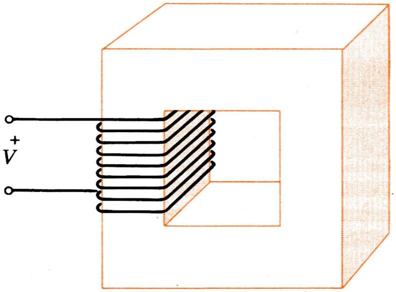


Figure 22.8 Calculation of core loss in an iron-core inductor.

Neglecting the *IR* drop, the induced voltage must be equal to the applied voltage, and by Eq. 22-3, the flux density must be

$$B_m = k\frac{V}{f}$$

where the constant k includes the number of turns and the core area. The ratio of the new B'_m to the old B_m is

$$\frac{B'_m}{B_m} = \frac{V'}{V}\cdot\frac{f}{f'} = \frac{150}{120}\cdot\frac{60}{50} = \frac{3}{2}$$

By Eq. 22-11 (assuming $n = 1.6$),

$$\frac{P'_h}{P_h} = \frac{f'}{f}\cdot\left(\frac{B'_m}{B_m}\right)^{1.6} = \left(\frac{5}{6}\right)\left(\frac{3}{2}\right)^{1.6} = 1.6$$

By Eq. 22-12,

$$\frac{P'_e}{P_e} = \left(\frac{f'}{f}\right)^2\left(\frac{B'_m}{B_m}\right)^2 = \left(\frac{5}{6}\right)^2\left(\frac{3}{2}\right)^2 = 1.56$$

The operation of electromagnetic devices can be greatly affected by relatively small changes in operating conditions. (How is I' related to I?)

Transformer Operation

The transformer is an electromagnetic energy converter whose operation can be explained in terms of the behavior of a magnetic circuit excited by an alternating current. In its most common form, the transformer consists of two (or more) multiturn coils wound on the same magnetic core but insulated from it. A changing voltage applied to the input or *primary* coil causes a changing current to flow thus creating a changing magnetic flux in the core. Because of the changing flux, voltage is induced in the output or *secondary* coil. No electrical connection between input and output is necessary; the transformer may be used to insulate one circuit from another while permitting an exchange of energy between them. Since only changing currents are transformed, an output circuit can be isolated from a direct-current component of the input.

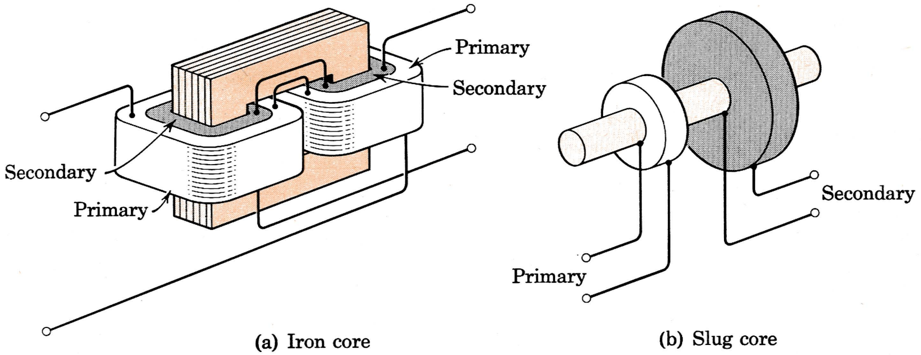
---

† See p. 392 of W. H. Timbie, V. Bush, and G. Hoadley, *Principles of Electrical Engineering*, 4th ed., John Wiley & Sons, New York, 1951.

By adjusting the ratio of the turns on the two coils, we can obtain a voltage "step up" or "step down." The same device can be used to obtain a current step up or step down. In addition to transforming voltage or current, a transformer may be used to transform impedance to obtain maximum power transfer through impedance matching. Because these functions are performed efficiently and precisely, the transformer is an important energy conversion device and careful study of its behavior is justified.

Construction

In the basic configuration, two coils are wound on a common core. For power applications in the 25- to 400-Hz frequency range, close coupling is desired and the coils are intimately wound on a highly permeable closed iron core (Fig. 22.9a).



(a) Iron core (b) Slug core

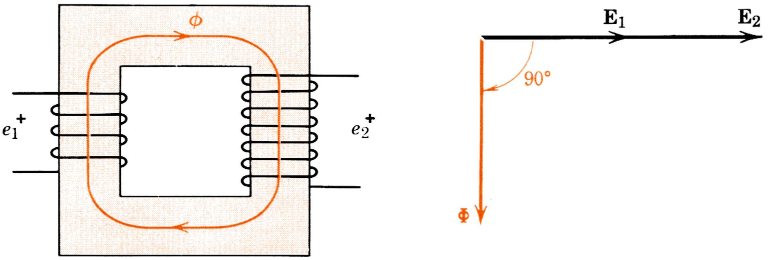
Figure 22.9 Two transformers.

Special design refinements make iron-core transformers useful over the audio-frequency range (20 to 20,000 Hz). For high frequencies (hundreds of kilohertz) and loose coupling, coils may be wound on a powdered iron "slug" (Fig. 22.9b) or with an air core.

To minimize resistance losses, the coils are usually wound with high-conductivity copper. In power and audio transformers, the cores are of high-permeability steel selected for low hysteresis loss and laminated to minimize eddy-current loss. The sum of hysteresis and eddy-current power is called *core loss* or *iron loss* in contrast to the *copper loss* due to the I^2R power in the windings.

Voltage Relations

In practice, the two coils are placed close together so that they link nearly the same flux. For clarity, in Fig. 22.10a the primary and secondary are shown on separate legs of the core, but it is assumed that both coils link the same flux. The voltage induced by the changing flux we call an *electromotive force* (emf) represented by the symbol e to distinguish it from the terminal voltage v.



(a) Schematic drawing **(b) Phasor diagram**

Figure 22.10 Transformer operation.

Assuming a sinusoidal variation in magnetic flux of the form $\phi = \Phi_m \sin \omega t$, the induced emfs are

$$e_1 = N_1 \frac{d\phi}{dt} = N_1 \omega \Phi_m \cos \omega t = \sqrt{2}\, E_1 \cos \omega t \tag{22-13a}$$

and

$$e_2 = N_2 \frac{d\phi}{dt} = N_2 \omega \Phi_m \cos \omega t = \sqrt{2}\, E_2 \cos \omega t \tag{22-13b}$$

where E_1 and E_2 are the rms values of the sinusoidal emfs. Therefore,

$$\frac{e_2}{e_1} = \frac{E_2}{E_1} = \frac{N_2}{N_1} \tag{22-14}$$

or the emf ratio is just equal to the *turn ratio.* The relation between emf and magnetic flux phasors is shown in Fig. 22.10b. Since $\sin \omega t = \cos (\omega t - 90°)$, the flux phasor lags behind the emf phasor by 90°.

In practical transformers, the terminal voltages differ only slightly from the induced emfs and the terminal voltage ratio is approximately equal to the turn ratio. As an illustration, there may be a transformer on a power pole on your rear property line to transform the voltage from an efficient transmission value of 4000 V to a safe working value of 240 V. (Approximately what turn ratio is required?)

Exciting Current

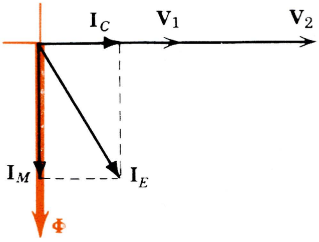


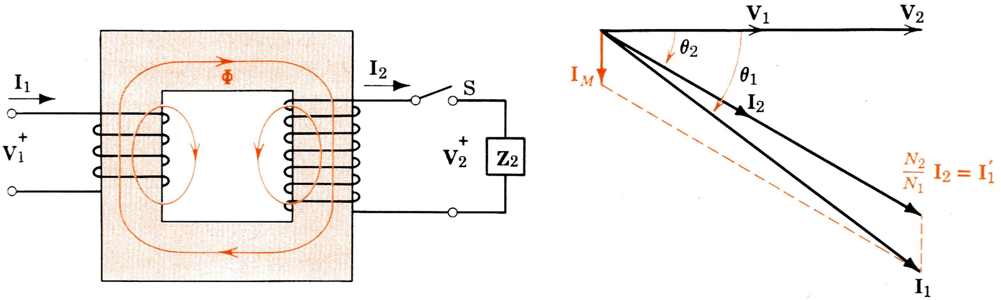
Figure 22.11 Components of the exciting current.

With the secondary open-circuited, the emf E_2 appears across the terminals, V_2 is just equal to E_2, and V_1 is approximately equal to E_1. The current that flows in the primary when the secondary is open-circuited is called the *exciting current*. As shown in Fig. 22.11, the exciting current $\mathbf{I}_E$ consists of two parts. One component, the *magnetizing current* $\mathbf{I}_M$, establishes the necessary flux to satisfy Eq. 22-13a and can be calculated when the properties and dimensions of the iron core are known. As expected, $\mathbf{I}_M$ is in phase with $\mathbf{\Phi}$ which it produces. The other component, the *core loss current* $\mathbf{I}_C$, represents the power dissipated in hysteresis and eddy-current loss. As expected, $\mathbf{I}_C$ is in phase with voltage $\mathbf{V}_1$ so that the product is power.

In an efficient transformer the core loss is small and $\mathbf{I}_E$ is approximately equal to $\mathbf{I}_M$. Strictly speaking, the magnetizing current is not sinusoidal for sinusoidal voltage and flux because of the nonlinear relation between $H = Ni/l$ and $B = \Phi/A$. However, for many transformers the magnetizing current is relatively small, and it is convenient to assume that i_M is sinusoidal and can be represented by phasor $\mathbf{I}_M$.

Current Relations

To develop the relation between primary and secondary currents under load conditions, let us assume (Fig. 22.12) that $\mathbf{V}_2 = \mathbf{E}_2$, $\mathbf{V}_1 = \mathbf{E}_1$, and $\mathbf{I}_E = \mathbf{I}_M$. When load $\mathbf{Z}_2$ is connected across the secondary by closing switch S, a current $\mathbf{I}_2 = \mathbf{V}_2/\mathbf{Z}_2$ flows. (Here θ_2 is assumed to be about 30° and the turn ratio $N_2/N_1 \cong 2$.) The load current $\mathbf{I}_2$ produces an mmf $N_2\mathbf{I}_2$ that *tends* to oppose



(a) Circuit (b) Phasor diagram

Figure 22.12 Current relations in a loaded transformer.

the magnetic flux which produces it. (Using the right-hand rule, check the directions of the mmfs produced by positive currents in the primary and secondary coils.) But the flux cannot change if $\mathbf{E}_1$ is to equal $\mathbf{V}_1$, and therefore additional primary current $\mathbf{I}_1'$ must flow. To maintain the core flux, the new net mmf must equal the mmf due to $\mathbf{I}_M$ alone or

$$N_1\mathbf{I}_M - N_2\mathbf{I}_2 + N_1\mathbf{I}_1' = N_1\mathbf{I}_M \tag{22-15}$$

This requires that

$$N_2\mathbf{I}_2 = N_1\mathbf{I}_1'$$

and

$$\frac{\mathbf{I}_1'}{\mathbf{I}_2} = \frac{N_2}{N_1} \tag{22-16}$$

We conclude that a secondary current $\mathbf{I}_2$ causes a component of primary current $\mathbf{I}_1'$; these currents are in phase and their ratio is just equal to the turn ratio. (The ratio I_2/I_1' is the *reciprocal* of the turn ratio.)

Strictly speaking, some of the mmf N_2I_2 goes to produce the leakage flux shown in Fig. 22.12a; also, all the flux produced by the primary does not link the secondary. In a well-designed transformer, however, the leakage flux is small and there is little error in Eqs. 22-14 and 22-16.

Because the transfer of the effect of a load across the output of a transformer to the source supplying the input is a difficult concept to grasp, let us describe the process again in different words. With the secondary open (the transformer unloaded), the input current is just the exciting current that establishes the necessary magnetic flux (so that $\mathbf{E}_1 \cong \mathbf{V}_1$) and supplies the core losses. The no-load power input is essentially the power dissipated in the core. For a given $\mathbf{V}_2$, the secondary current that flows when a load is connected is determined in magnitude and angle by $\mathbf{Z}_2$. To maintain the magnetic flux at the required value, the mmf $N_2\mathbf{I}_2$ produced by the secondary current must be offset by an equal and opposite mmf $N_1\mathbf{I}_1'$ produced by additional primary current $\mathbf{I}_1'$. To the source supplying the transformer, the increased primary current represents an increase in apparent power. The total power input to the transformer is $P_1 = V_1I_1 \cos\theta_1$, where $\mathbf{I}_1$ is the phasor sum of the exciting current and the load current transferred to the primary. (See Example 4.)

Linear Circuit Models

In deriving the voltage and current relations for a transformer and in solving Example 4, we made many simplifying assumptions. With just a little more effort we can derive a model that provides much greater precision. Furthermore, in spite of the nonlinearity of the magnetic circuit and the vagaries of the

Example 4

The laminated silicon steel core of a transformer (Fig. 22.13) has a mean length of 0.6 m and a cross-section of 0.005 m^2. There are 150 turns on the primary (actually split between two "legs" of the core as in Fig. 22.9) and 450 turns on the secondary. The input is 200 V rms at 60 Hz. Estimate the primary current with the secondary open and with a resistance load of 120 Ω connected across the secondary.

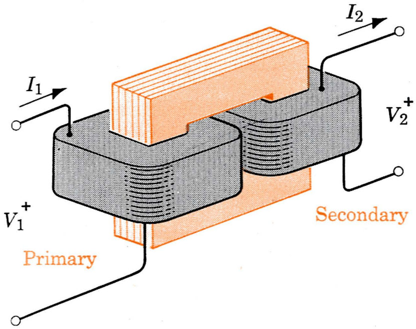


Figure 22.13 Transformer calculations.

The necessary maximum flux is, by Eq. 22-13a,

$$\Phi_m = \frac{\sqrt{2}E_1}{N_1\omega} = \frac{\sqrt{2} \times 200}{150 \times 2\pi \times 60} = 5 \times 10^{-3}\ \text{Wb}$$

The flux density is

$$B_m = \frac{\Phi_m}{A} = \frac{5 \times 10^{-3}}{5 \times 10^{-3}} = 1\ \text{T}$$

For silicon steel (Fig. 21.11) at this density, $H = 200$ A-t/m. The magnetizing current (rms value) is

$$I_M = \frac{1}{\sqrt{2}}\frac{\mathcal{F}}{N_1} = \frac{Hl}{\sqrt{2}N_1} = \frac{200 \times 0.6}{\sqrt{2} \times 150} \cong 0.6\ \text{A}$$

Since silicon steel has low hysteresis loss and the laminated core has low eddy-current loss, as an estimate we assume that $I_E = I_M = 0.6$ A.

Assuming that the voltage ratio is equal to the turn ratio,

$$V_2 = \frac{N_2}{N_1}V_1 = \frac{450}{150} \times 200 = 600\ \text{V}$$

With the resistance load connected,

$$I_2 = \frac{V_2}{R_2} = \frac{600\ \text{V}}{120\ \Omega} = 5\ \text{A}$$

The component of primary current due to the load is

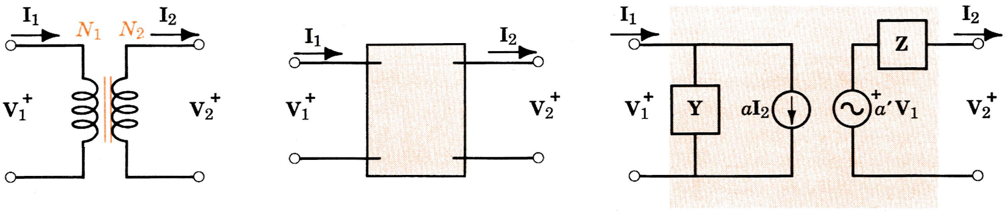
$$I_1' = \frac{N_2}{N_1}I_2 = \frac{450}{150} \times 5 = 15\ \text{A}$$

Since the exciting current is small and is to be added at nearly right angles to I_1' (see Fig. 22.12b), it is neglected and the primary current with load ≅ 15 A.

hysteresis effect, precise results over a wide range of operating conditions can be obtained using a linear model. Finally, and this is important to engineers, the model parameters for a given transformer can be determined by two simple tests that can be performed with readily available instruments.

The Transformer as a Two-Port

In deriving a linear circuit model, we use the approach that worked well with transistors and incorporate our understanding of transformer operation. In wiring diagrams, an iron-core transformer is represented by the schematic symbol in Fig. 22.14a. In the two-port representation of Fig. 22.14b, we follow convention and show $\mathbf{I}_2$ as the current *out* because the transformer is considered to be a source supplying a load connected to the output port. Since transformers work only with changing currents, usually sinusoidal, we think in terms of phasor quantities and complex immittances.



(a) Schematic symbol (b) Two-port model (c) Hybrid parameter model

Figure 22.14 Transformer representation.

In deriving a circuit model, we can choose any two independent variables and the choice determines the nature of the model parameters. Let us choose $\mathbf{V}_1$ and $\mathbf{I}_2$ (see Eq. 17-11) and write

$$\begin{cases} \mathbf{I}_1 = f_1(\mathbf{V}_1, \mathbf{I}_2) = \mathbf{Y}\mathbf{V}_1 + a\mathbf{I}_2 & \text{(22-17)} \\ \mathbf{V}_2 = f_2(\mathbf{V}_1, \mathbf{I}_2) = a'\mathbf{V}_1 - \mathbf{Z}\mathbf{I}_2 & \text{(22-18)} \end{cases}$$

In these relations we recognize four hybrid parameters that are constant in a linear model. The first equation indicates that $\mathbf{I}_1$ has two components and suggests a parallel circuit. The second equation indicates that $\mathbf{V}_2$ has two components and suggests a series circuit. (The minus sign arises because $\mathbf{I}_2$ is defined as the current *out*.) The result is shown in Fig. 22.14c; note that there is no electrical connection between input and output.

What is the physical nature of each of the four parameters? With the secondary open circuited, $\mathbf{I}_2 = 0$ and $\mathbf{I}_1$ is just equal to $\mathbf{I}_E$, the exciting current. By Eq. 22-17,

$$\mathbf{Y} = \frac{\mathbf{I}_E}{\mathbf{V}_1} = \frac{I_C}{V_1} - j\frac{I_M}{V_1} = G + jB \qquad \text{(22-19)}$$

where I_C = the core-loss current (Fig. 22.15),
I_M = the magnetizing current,
G = an admittance accounting for power loss, and
B = an inductive (negative) susceptance accounting for energy storage.

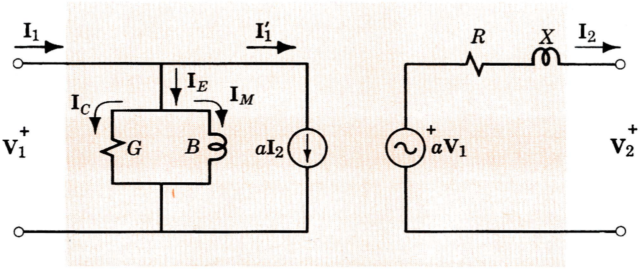


Figure 22.15 Hybrid parameter model of a transformer.

Since the open-circuit secondary voltage is approximately the primary voltage times the turn ratio, Eq. 22-18 indicates that

$$a' = \frac{N_2}{N_1} \tag{22-20}$$

With a current $\mathbf{I}_2$ flowing in the secondary, we expect (Eq. 22-16) a component $\mathbf{I}_1'$ equal to $(N_2/N_1)\mathbf{I}_2$ to appear in the primary. Therefore,

$$a = \frac{N_2}{N_1} = a' \tag{22-21}$$

With the transformer loaded, i.e., $\mathbf{I}_2 \neq 0$, the power lost in the resistance of the windings and the energy stored in the leakage fields (Fig. 22.12a) become appreciable. These effects are accounted for by the impedance

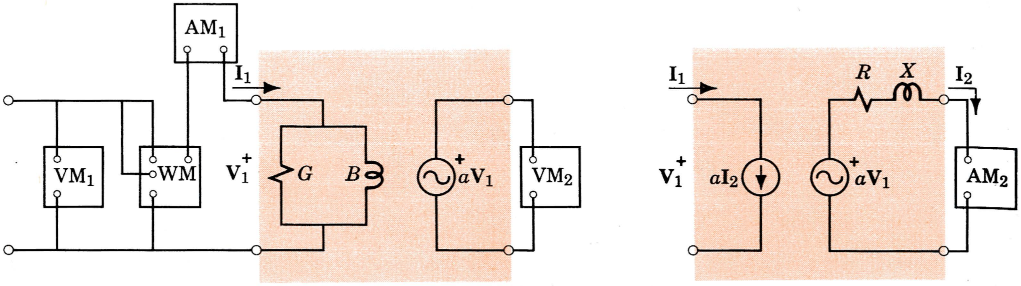
$$\mathbf{Z} = R + jX = R + j\omega L \tag{22-22}$$

where R = an equivalent resistance including the effects of both windings,
X = an equivalent reactance, and
L = an equivalent leakage inductance.

Because of $\mathbf{Z}$, the output voltage changes with variations in load current.

Open- and Short-Circuit Tests

The parameters in the linear circuit model are easily and accurately determined by an ingenious laboratory procedure. With the secondary open circuited, $\mathbf{I}_2 = 0$, $\mathbf{I}_1 = \mathbf{I}_E$, and the transformer model is as shown in Fig. 22.16a. Voltmeter VM_2 has a very high resistance and appears to be an open circuit. Wattmeter WM measures the total power input, and VM_1 and AM_1 measure the primary voltage and exciting current, respectively.



(a) Open–circuit test (b) Short–circuit test

Figure 22.16 Instrumentation for determining model parameters.

In this *open-circuit test* the exciting current flows in the primary windings. However, I_E is small and since power varies as the square of the current the power lost in the windings is negligible and WM indicates only the core loss. Therefore, the parameters are

$$G = \frac{P_{OC}}{V_{1O}^2} \qquad Y = \frac{I_{1O}}{V_{1O}} \tag{22-23}$$

$$B = -\sqrt{Y^2 - G^2} \qquad a = \frac{V_{2O}}{V_{1O}} \tag{22-24}$$

Since the core loss is dependent on B_m and f (see Example 3), and therefore dependent on V_1, the open-circuit test is performed at rated voltage and frequency. (Transformer ratings are discussed on p. 589.)

In contrast, the *short-circuit test* is performed at rated current. With the secondary short-circuited (ammeter AM_2 has a very low resistance and appears to be a short circuit), the transformer model is as shown in Fig. 22.16b. Since $\mathbf{I}_2$ is limited only by the small internal impedance $R + jX$, the primary voltage required for rated current is very small. At this low voltage B_m is very small, the core loss is very small, and the exciting admittance $\mathbf{Y}$ is omitted from the model.

With the instruments of Fig. 22.16a applied to the short-circuited model, wattmeter WM indicates only the copper loss. Since $P_{SC} = I_{2S}^2 R = (I_{1S}/a)^2 R$,

$$R = \frac{a^2 P_{SC}}{I_{1S}^2} \tag{22-25}$$

Then

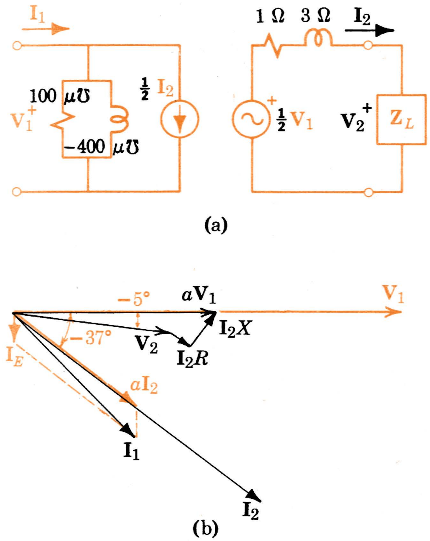
$$Z = \frac{aV_{1S}}{I_{2S}} = \frac{aV_{1S}}{I_{1S}/a} = a^2 \frac{V_{1S}}{I_{1S}} \tag{22-26}$$

and

$$X = \sqrt{Z^2 - R^2} = \sqrt{(a^2 V_{1S}/I_{1S})^2 - R^2} \tag{22-27}$$

Example 5

The primary of a transformer is rated at 10 A and 1000 V. On open circuit, instruments connected as in Fig. 22.16a indicate: $V_1 = 1000$ V, $V_2 = 500$ V; $I_1 = 0.42$ A, $P_{OC} = 100$ W. On short circuit, the readings are: $I_1 = 10$ A, $V_1 = 126$ V, $P_{SC} = 400$ W. Determine the hybrid parameters, predict the output voltage across a load impedance $\mathbf{Z}_L = 19 + j12\ \Omega$, and draw a phasor diagram.



$\mathbf{I}_1$
1 Ω
3 Ω
$\mathbf{I}_2$
100 µ℧
$\frac{1}{2}\mathbf{I}_2$
$\mathbf{V}_1$
−400 µ℧
$\frac{1}{2}\mathbf{V}_1$
$\mathbf{V}_2$
$\mathbf{Z}_L$
−5°
$a\mathbf{V}_1$
$\mathbf{V}_1$
$\mathbf{I}_2X$
−37°
$\mathbf{V}_2$
$\mathbf{I}_2R$
$\mathbf{I}_E$
$a\mathbf{I}_2$
$\mathbf{I}_1$
$\mathbf{I}_2$

(a)

(b)

Figure 22.17 Transformer performance.

The parameters are determined from the test data. From open-circuit data,

$$G = \frac{P_{OC}}{V_{1O}^2} = \frac{100}{(1000)^2} = 100 \times 10^{-6} = 100\ \mu℧$$

$$Y = \frac{I_{1O}}{V_{1O}} = \frac{0.42}{1000} = 420\ \mu℧$$

$$\begin{aligned} B &= -\sqrt{Y^2 - G^2} \\ &= -\sqrt{(420)^2 - (100)^2} \times 10^{-6} \cong -400\ \mu℧ \end{aligned}$$

$$a = V_{2O}/V_{1O} = 500/1000 = 1/2$$

From short-circuit data,

$$R = \frac{a^2 P_{SC}}{I_{1S}^2} = \frac{400}{4 \times 10^2} = 1\ \Omega$$

$$Z = a^2 \frac{V_{1S}}{I_{1S}} = \frac{126}{4 \times 10} = 3.15\ \Omega$$

$$X = \sqrt{Z^2 - R^2} = \sqrt{(3.15)^2 - (1)^2} \cong 3\ \Omega$$

The labeled circuit model and specified load are shown in Fig. 22.17a. The secondary current under load is

$$\begin{aligned} \mathbf{I}_2 &= \frac{\frac{1}{2}\mathbf{V}_1}{\mathbf{Z} + \mathbf{Z}_L} = \frac{\frac{1}{2} \times 1000\underline{/0^\circ}}{(1 + j3) + (19 + j12)} \\ &= 20\underline{/-37^\circ}\ \text{A} \end{aligned}$$

and the output voltage is

$$\mathbf{V}_2 = \mathbf{I}_2\mathbf{Z}_L = (20\underline{/-37^\circ})(22.5\underline{/+32^\circ}) = 450\underline{/-5^\circ}\ \text{V}$$

Note that under this load, the terminal voltage ratio is less than the turn ratio.

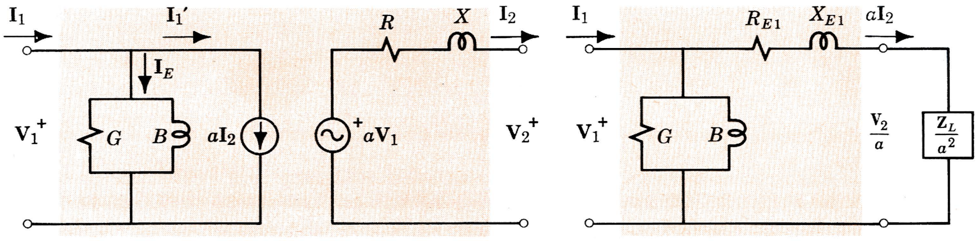
The complete phasor diagram is shown in Fig. 22.17b with $I_E = I_{1O}$ exaggerated in scale. Note that $a\mathbf{V}_1$ is equal to $\mathbf{V}_2 + \mathbf{I}_2\mathbf{Z}$.

Ammeter AM_2 is not essential, but the ratio I_{1S}/I_{2S} provides a check on the value of a determined from V_{20}/V_{10} in the open-circuit test.

There are two main reasons for representing a device by a linear model. First, the model permits general analysis and the drawing of general conclusions. Second, once the parameters are determined, the behavior of the device under various operating conditions can be predicted as in Example 5.

Other Models

The hybrid parameter model of Fig. 22.18a is convenient to work with; the parameters are easily determined and calculations are straightforward. Since an impedance $\mathbf{Z}$ in the secondary circuit has the same effect as an impedance $\mathbf{Z}/a^2$ in the primary, the model can be converted to the alternate form of Fig. 22.18b. Here



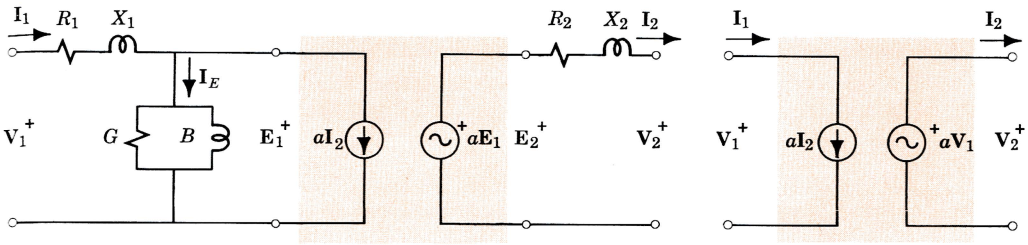
(a) Hybrid parameters (b) Equivalent parameters

Figure 22.18 A model with parameter values referred to the primary.

$$\frac{R}{a^2} = R_{E1} \quad \text{and} \quad \frac{X}{a^2} = X_{E1} \tag{22-28}$$

where R_{E1} and X_{E1} are equivalent resistance and equivalent reactance *referred to the primary*. Similarly, V_2/a is the secondary voltage referred to the primary and $\mathbf{Z}_L/a^2$ is the load impedance referred to the primary. By using these parameters, a transformer problem is reduced to a simple parallel circuit problem.

A defect of the model of Fig. 22.18 is that it indicates that the exciting current is independent of the load current; this is not quite true. The induced emf $\mathbf{E}_1$ differs from the primary voltage $\mathbf{V}_1$ by the amount of the voltage drop across the resistance and leakage reactance of the primary winding. As the load current increases, the primary current increases, the voltage drop increases, and the necessary emf $\mathbf{E}_1$ decreases slightly. At high load currents, $\mathbf{E}_1$ is less than at no load, the necessary flux $\mathbf{\Phi}$ is less, and therefore $\mathbf{I}_E$ is less. If the portion of $\mathbf{Z}$ due to the primary resistance and leakage reactance is placed as shown in the conventional model of Fig. 22.19a, the reduction of $\mathbf{E}_1$ and $\mathbf{I}_E$ with increasing $\mathbf{I}_1$ is accounted for automatically. A second improvement is that here $a = \mathbf{E}_2/\mathbf{E}_1$, which is just equal to the turn ratio and is constant. Because of practical design considerations, in commercial transformers the resistances and leakage reactances



(a) A precise circuit model

(b) An ideal transformer

Figure 22.19 Conventional transformer models.

divide so that $R_1 \cong R_2/a^2$ and $X_1 \cong X_2/a^2$ (see Problem 6). In performing the short-circuit test, the impedance seen at the primary terminals is

$$\frac{\mathbf{V}_{1S}}{\mathbf{I}_{1S}} = \left(R_1 + \frac{R_2}{a^2}\right) + j\left(X_1 + \frac{X_2}{a^2}\right) \tag{22-29}$$

Example 6

Using the data of Example 5, determine the parameters for the more precise model of Fig. 22.19.

The values of G, B, and a are as previously calculated. Assuming the previously calculated resistance R and leakage reactance X divide in the usual way,

$$R_2 = (\tfrac{1}{2})R = \tfrac{1}{2}(1) = 0.5\,\Omega$$

$$R_1 = \frac{1}{a^2}R_2 = 4R_2 = 4 \times 0.5 = 2.0\,\Omega$$

$$X_2 = \tfrac{1}{2}X = \tfrac{1}{2}(3) = 1.5\,\Omega$$

$$X_1 = \frac{1}{a^2}X_2 = 4X_2 = 4 \times 1.5 = 6\,\Omega$$

The portion within the shaded area in Fig. 22.19b is called an *ideal transformer* because it provides for the transformation of voltage, current, and impedance without any of the imperfections of a real transformer. A real transformer approaches the ideal as the conductivity of the windings approaches infinity, the leakage flux approaches zero, the permeability of the core approaches infinity, and the core loss approaches zero. The ideal transformer is sometimes represented by adding the label "ideal" to the symbol in Fig. 22.14a, but its character is more clearly indicated by the symbolic representation in Fig. 22.19b.

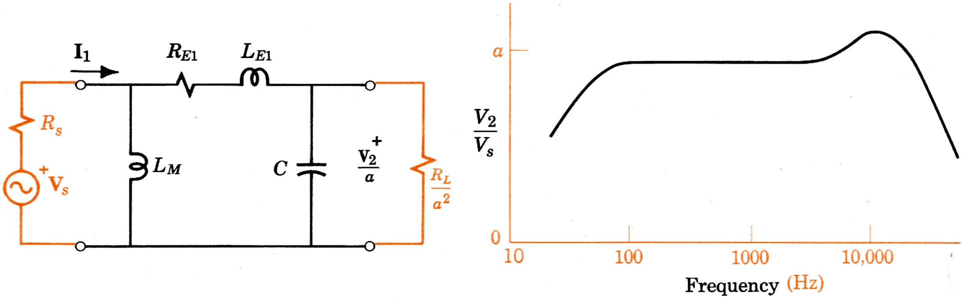
Selecting a Model

The ideal transformer and the precise model represent the extremes of complexity ordinarily employed in power transformer analysis. In a given situation, the best model to use is the one that gives the required degree of precision with a minimum of effort; the hybrid parameter model of Fig. 22.18a is one useful compromise. If the transformer is heavily loaded, the exciting admittance Y may be of negligible importance. If voltage variation is of greater interest than power dissipation, the resistive elements (R and G) can be neglected and only the relatively larger reactive elements (X and B) retained. If core loss is unimportant, let $I_E = I_M$ or $G = 0$. (See Exercise 24.)

Models for Communications Transformers

Power transformers are ordinarily operated at fixed frequencies and the circuit model reactances are constant. The input and output transformers used in audiofrequency applications (for example, in connecting a power transistor to a loud speaker) must handle signals from, say, 20 to 15,000 Hz. Because of the wide variation in reactances, the circuit model applicable at low frequencies is quite different from that appropriate at high frequencies. In addition the *distributed capacitance* of the windings, unimportant in most power transformer applications, becomes significant at high frequencies.

A convenient model for an output transformer is shown in Fig. 22.20a. This



(a) Circuit model

(b) Frequency response

Figure 22.20 An audiofrequency transformer.

is derived from the model of Fig. 22.18b by neglecting G, replacing frequency-dependent B by L_M, and including a lumped shunt capacitance C to represent all the distributed capacitive effects. Laboratory measurements on the actual transformer indicate a response curve like that shown in Fig. 22.20b.

At moderate frequencies, corresponding to the midfrequency range of an amplifier, the admittances of the shunt paths represented by L_M and C and the reactance of the series element L_{E1} are negligible. If R_S and R_{E1} are small, the mid-frequency voltage gain is just the turn ratio a. At low frequencies, the admittance $Y = 1/\omega L_M$ of the magnetizing circuit is high, part of the input current flows through this shunt path, and the output is reduced. At high frequencies, the effect of L_M is negligible, but the effects of L_{E1} and C are important.

At the series resonant frequency of L_{E1} and C the voltage gain may be greater than a, but at higher frequencies the gain drops sharply. For quantitative analysis, the general model of Fig. 22.20a can be replaced by three simplified models that are applicable over restricted frequency ranges (see Exercise 25).

Transformer Performance

For power applications the important external characteristics are voltage ratio, output power, efficiency, and the voltage variation with load. These may be obtained from manufacturer's specifications, from experimental measurements, or from calculations based on a circuit model.

Nameplate Ratings

The manufacturer of an electrical machine usually indicates on the *nameplate* the normal operating conditions.† A typical nameplate might read: "Transformer, 4400:220 V, 60 Hz, 10 kVA." The design voltages of the two windings are 4400 V and 220 V rms and the turn ratio is 1:20; either side may be the primary. At a frequency of 60 Hz, the design voltages bring operation near the knee of the magnetization curve and exciting current and core losses are not excessive. Using either side as a secondary, the rated *output* of 10 kVA ("full" load) can be maintained continuously without excessive heating (and the consequent deterioration of the winding insulation). Because the heating is dependent on the square of the current, the output is rated in apparent power (kVA) rather than in power (kW). Supplying a zero power-factor load, a transformer can be operating at rated output while delivering zero power.

The cross-section of the iron core is determined by the operating voltage and frequency. The cross-section of the copper conductor is determined by the operating current. Knowing the effect of these factors, an engineer can *rerate* a device under changed operating conditions. (For example, what should be the 50-Hz rating of the transformer whose nameplate data are given in the preceding paragraph?)

Efficiency

Efficiency is by definition the ratio of output power to input power. A convenient form is

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{\text{output}}{\text{output} + \text{losses}} = 1 - \frac{\text{losses}}{\text{input}} \qquad (22\text{-}30)$$

Using the hybrid parameter model of Fig. 22.15, the output is $V_2 I_2 \cos\theta_2$ and the copper loss is $I_2^2 R$. The iron loss is independent of load current and is given by $V_1^2 G$ which is just equal to P_{OC} (Eq. 22-23).

By increasing the amount of copper, the I^2R loss is reduced; by increasing the amount (or quality) of iron, the eddy-current and hysteresis losses are reduced. The optimum design is based on economic considerations as well as technical factors. A *distribution* transformer that is always connected to the line should be designed with low iron losses. A *power* transformer that is only connected when in actual use usually has higher iron losses and lower copper losses than a comparable distribution transformer. It can be shown that for many engineering devices the maximum efficiency occurs when the variable losses are equal to the fixed losses (see Problem 7). In an *instrument* transformer that converts high voltages or high currents to lower values that are more easily measured, the important characteristic is a constant ratio of transformation.

---

† See the section on Root-Mean-Square Rating on p. 80.

Voltage Regulation

To hold the speed of a motor or the voltage of a generator constant under varying load we employ a *regulator*, and the variation in speed, or voltage, is

Example 7

The 10-kVA, 60-Hz, 1000/500-V transformer of Example 5 supplies a 0.5 leading pf load. Predict the voltage regulation and the full-load efficiency.

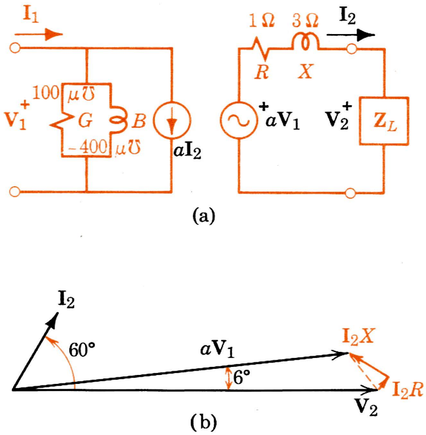


Figure 22.21 Voltage regulation and efficiency.

The model of Fig. 22.17 is repeated in Fig. 22.21. The internal impedance of the transformer is

$$\mathbf{Z} = R + jX = 1 + j3 = 3.16\underline{/+71.5^\circ}\ \Omega$$

The output is rated at 500 V and 10 kVA; therefore, $V_2 = 500$ V and $I_2 = 10{,}000/500 = 20$A.

Under full load at 0.5 leading pf, the voltage drop across the internal impedance of the transformer is

$$\mathbf{V}_Z = \mathbf{I}_2\mathbf{Z} = 20\underline{/60^\circ} \times 3.16\underline{/71.5^\circ} = 63.2\underline{/131.5^\circ}\ \text{V}$$

The phasor diagram (Fig. 22.21b) is drawn with $\mathbf{V}_2$ as a reference; it indicates that the no-load voltage $a\mathbf{V}_1$ is

$$\mathbf{V}_2 + \mathbf{I}_2\mathbf{Z} = (500 + j0) + (-41 + j48) \cong 462\underline{/6^\circ}\ \text{V}$$

The voltage regulation (Eq. 22-31) is

$$\text{VR} = \frac{aV_1 - V_2}{V_2} = \frac{462 - 500}{500} = -0.076 \text{ or } -7.6\%$$

The negative VR means a rise in output voltage due to leading pf current in the leakage reactance X.

To determine efficiency, we note that the iron loss is given by $P_{OC} = 100$ W and the copper loss is $I_2^2R = (20)^2(1) = 400$ W. The output is $V_2I_2\cos\theta = 500 \times 20 \times 0.5 = 5000$ W. By Eq. 22-30,

$$\text{Efficiency} = \frac{5000}{5000 + 100 + 400} \cong 0.91 \text{ or } 91\%$$

Even at this poor pf the efficiency is quite high.

expressed by the *regulation.* By definition the regulation is

$$\text{Regulation} = \frac{\text{no-load value} - \text{full-load value}}{\text{full-load value}} \tag{22-31}$$

The voltage regulation of a transformer, then, is the change in output voltage (produced by a change in load current from zero to rated value) divided by the rated voltage. The regulation may be either positive or negative and is usually expressed in percent. (See Example 7.)

The voltage regulation is defined for voltage *magnitudes,* but usually it is calculated from a phasor diagram based on a circuit model. The primary cause of voltage drop with load is leakage reactance and if high precision is not required, a simplified model can be used. Note that the regulation is positive if the voltage decreases with increasing load.

SUMMARY

- With ac excitation of a coil, the magnetic flux in the core is just that required to induce a voltage approximately equal to the applied voltage.
  For a linear magnetic circuit, the induced voltage is

  $$v = N\frac{d\phi}{dt} = \frac{d\lambda}{dt} = L\frac{di}{dt}$$

  For a flux $\phi = \Phi_m \sin 2\pi ft$, $V(\text{rms}) = 4.44Nf\Phi_m$

- The energy density (energy/unit volume) in a magnetic field is

  $$W_V = \int_0^B H\,dB$$

  In many problems all the energy stored can be assumed to be in the air gap.

- In ferromagnetic materials subjected to alternating mmfs, the iron losses are due to hysteresis and eddy currents, and

  $$P_i = P_h + P_e = K_h f B_m^n + K_e f^2 B_m^2$$

- Transformers convert ac power from one voltage, current, or impedance level to another level. In an ideal transformer,

  $$\frac{V_2}{V_1} = \frac{N_2}{N_1} = a \qquad \frac{I_2}{I_1} = \frac{N_1}{N_2} \qquad \frac{Z_2}{Z_1} = \left(\frac{N_2}{N_1}\right)^2$$

  In a practical transformer these relations are modified because of winding resistance, leakage reactance, and exciting current.
  The exciting current includes magnetizing and core loss components.

- Transformer performance can be predicted on the basis of a linear circuit model of appropriate complexity and precision.
  Open- and short-circuit tests determine parameter values.
  Distributed capacitance is important at high frequencies.
- Transformers can operate at rated voltage and current continuously.

$$\text{Efficiency} = \frac{\text{output}}{\text{output} + \text{iron loss} + \text{copper loss}}$$

$$\text{Regulation} = \frac{\text{no-load value} - \text{full-load value}}{\text{full-load value}}$$

REVIEW QUESTIONS

1. Why were iron wires used in the Déri transformers?
2. What are the advantages of a closed core over an open core?
3. What determines the magnetic flux in a core excited by a direct voltage? An alternating voltage?
4. What is the effect on inductance of doubling the number of turns on a given core?
5. Describe an experimental method for measuring the energy stored in a given magnetic core.
6. Under what circumstances can all the energy stored in a magnetic circuit be assumed to be in the air gap?
7. What are eddy currents? How are they minimized?
8. Sketch a core with two windings, one connected by means of switch S_1 to a voltage source V_1 and the other connected by means of switch S_2 to a resistance R_2.
   (a) Explain the electric and magnetic behavior when S_1 is closed.
   (b) Explain the behavior when S_2 is then closed.
9. List four functions of a transformer.
10. How is induced emf different from terminal voltage?
11. What assumptions are made in deriving the model in Fig. 22.14b?
12. Compare the hybrid parameter models of the transistor and the transformer.
13. Can a 240:120-V transformer be used to step up 120 V to 240 V? To step up 240 V to 480 V?
14. What is the distinction between exciting and magnetizing currents?
15. Outline the procedure and draw the circuit connections in the experimental determination of circuit model parameters.
16. Draw from memory three circuit models of increasing precision.
17. What is meant by "rated kVA" of a transformer? Why not "rated kW"?
18. Define the speed regulation of an electric motor driving a saw.

EXERCISES

1. In Example 6 of Chapter 21, the 3-A current is provided from a battery through a switch. If the switch is suddenly opened and the flux is reduced to zero in 10 μs, predict the voltage across the switch.
2. A current i in the N-turn coil of Fig. 22.22 establishes magnetic fluxes ϕ_1 and ϕ_2 in time t. Predict the average voltage V appearing across terminals ab.

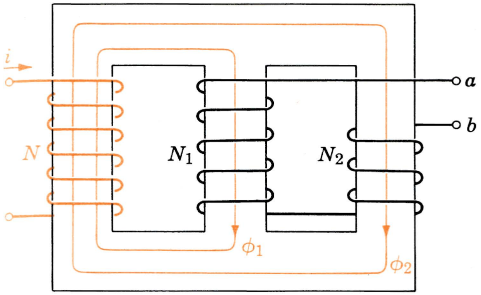


Figure 22.22

3. Compute the inductances of the toroidal coil of Example 2 in Chapter 21.
4. An N-turn inductance coil is wound on a high permeability core of cross-section A and length l_i with an air gap length l_a.
   (a) Stating any assumptions, derive an approximate expression for the inductance.
   (b) For $N = 800$ t, $A = 1\ \text{cm}^2$, $l_i = 10$ cm, $l_a = 1$ mm, and $I = 10$ mA, predict the inductance.
   (c) If the core of part (b) is cast steel and $I = 2$ A, check your assumptions and predict the inductance.
5. Sketch hysteresis loops that would be "ideal" for a transformer and for a computer memory element.
6. An inductance coil for 120-V, 60-Hz operation is to be wound on an Armco iron core 4 cm × 3 cm in cross-section and 30 cm in mean length. The maximum allowable flux density in the core ($\mu_r \cong 4000$) is 1.2 T.
   (a) How many turns are required?
   (b) What exciting current flows?
7. (a) Repeat Exercise 6(b) for the case of a 3-mm air gap in the core with the same N.
   (b) Compare the results for the two cases and draw a conclusion.
8. The hysteresis loss is 200 W and the eddy-current loss is 200 W when an iron-core inductance coil is operated at 120 V and 60 Hz. Estimate the losses for operation at: (a) 150 V and 60 Hz, (b) 120 V and 50 Hz, (c) 150 V and 72 Hz.
9. Redesign the coil of Exercise 8 (i.e., specify the necessary change in the number of turns) to reduce the total core loss to 300 W at 120 V and 60 Hz.
10. Locate a transformer of any type or size and examine it carefully. Draw a sketch with approximate dimensions and list the purpose, manufacturer, approximate weight, and voltage and current specifications if available.
11. The input to a power supply transformer (300 primary turns) is to be 120 V at 60 Hz. One secondary provides 120 mA at 500 V and another provides 2 A at 6 V (at unity pf).
    (a) Estimate the number of turns required on the secondaries.
    (b) Estimate the full-load primary current.
12. The core in Fig. 21.17 is made of silicon sheet steel with an air gap 2 mm long; $a = 4$ cm, $c = 5$ cm, and $b = d = 20$ cm. The 300-turn coil has a resistance of 2 Ω. Estimate the current for:
    (a) A steady magnetic flux of 2.5 mWb across the air gap.
    (b) An applied voltage of 120 V dc.
    (c) An applied voltage of 120 V at 60 Hz.
    (d) An applied voltage of 120 V at 40 Hz.
13. In a "clamp-on" ammeter, the core (similar to Fig. 21.17) is hinged so that it can be opened at the air gap and closed around any conductor carrying a current. The coil ($N = 20$ turns) supplies current to an ammeter. Sketch the connection and estimate the current flowing in the encircled conductor when the ammeter reads 1.5 A. Explain your reasoning.
14. A 500-W resistive load at 125 V is to be supplied from a 500-V, 60-Hz source using a transformer wound on an Armco iron core 10 cm² in cross-section and 35 cm long.
    (a) If the maximum allowable flux density is 1.2 T, approximately how many primary and secondary turns are required?
    (b) Approximately what is the magnetizing current?
    (c) Approximately what are the full-load primary and secondary currents?

15. In Fig. 22.23, $v = V_m \cos 2000t$ V, $R_s = 100\,\Omega$, $L = 10$ mH, $C = 1\,\mu$F, and $R = 25\,\Omega$. Design (i.e., specify the turn ratios $N_2/N_1 = a$ and b for) the transformers to provide maximum power transfer to the load R.

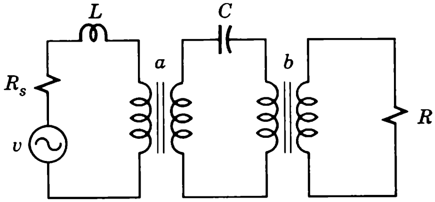


Figure 22.23

16. A 250:1000-V, 60-Hz transformer is tested. Open-circuit data are: $V_1 = 250$ V, $I_1 = 0.51$ A, $W_1 = 31.25$ W, $V_2 = 1000$ V. Short-circuit data are: $V_1 = 7.9$ V, $I_1 = 20$ A, $W_1 = 50$ W. Draw and determine the values for a hybrid parameter model.

17. The transformer of Exercise 16 is turned around and tested with power input to the 1000-V windings.
- (a) Predict the instrument readings for open- and short-circuit tests.
- (b) Draw and determine values for the hybrid parameter model of the transformer in this configuration.

18. The high-voltage winding of the transformer in Exercise 16 is connected to a 200-Ω resistance. Determine:
- (a) The actual primary voltage required to develop 1000 V across the resistance.
- (b) The actual primary current and show all voltages and currents on a phasor diagram.

19. Repeat Exercise 18 for a load impedance $\mathbf{Z}_L = 120 + j160\,\Omega$.

20. A transformer is labeled: 40 kVA, 4000:400 V, 60 Hz. The no-load primary current is 1 A. With the secondary short-circuited, the input to the primary is 1500 W at 250 V and 10 A.
- (a) Draw an appropriate circuit model, *stating* any necessary assumption.
- (b) Predict the primary current when a load $\mathbf{Z}_L = 3.85 - j0.2\,\Omega$ is connected across the 400-V secondary.
- (c) Repeat part b using the ideal transformer model and draw a conclusion.

21. For a certain 50-kVA, 400:2000-V, 60-Hz transformer, the model parameters (Fig. 22.19) are: $R_1 = 0.02\,\Omega$, $X_1 = 0.06\,\Omega$, $G = 2$ m℧, $B = -6$ m℧, $R_2 = 0.5\,\Omega$, and $X_2 = 1.5\,\Omega$.
- (a) Simplify the circuit model by neglecting the core loss, neglecting the voltage drop across $R_1 + jX_1$ due to the exciting current, and referring all quantities to the primary.
- (b) Using the simplified model, predict the no-load current and the primary current with a load $\mathbf{Z}_L = 100 + j75\,\Omega$.
- (c) For $\mathbf{Z}_L = 100 + j75\,\Omega$, predict the efficiency $= P_L/P_I$.

22. For the transformer of Exercise 21,
- (a) Simplify the circuit model as in part (a) by referring all quantities to the *secondary*.
- (b) Repeat parts (b) and (c) for $\mathbf{Z}_L = 59 - j48\,\Omega$.

23. A transformer is to be used to couple a 10-Ω loudspeaker to a transistor amplifier with a midfrequency output resistance of 4000 Ω. Specify the turn ratio and express the power gain (in decibels) obtained by using the transformer instead of connecting the loudspeaker directly across the amplifier.

24. Four 20-W 16-Ω loudspeakers connected in parallel are to be supplied by an amplifier with a 500-Ω output resistance. Specify the turn ratio of the output transformer and the output current of the amplifier.

25. Represent the transformer of Exercise 23 at low, moderate, and high frequencies by appropriately simplified circuit models.

26. The nameplate of a transformer with four separate coils reads: 10 kVA, 60 Hz, 460/230:230/115 V.
- (a) Interpret this nameplate and sketch the connection to provide a step down from 460 to 115 V.
- (b) What are the current ratings of the individual coils?

27. At a hydroelectric power plant the main transformer is a bank of three 66,667-kVA, 18,000:220,000-V units connected in Δ which steps up the generated voltage for transmission at 220 kV. Sketch the transformer connection and calculate the current ratings of the primary and secondary windings.

28. For a 300-kVA, 12,000:4000-V, 60-Hz transformer, open-circuit test readings are: 4000 V, 1.1 A, 2000 W, and short-circuit test readings are: 300 V, 25 A, and 2800 W.
(a) On which "side" of the transformer were the O–C and S–C tests made?
(b) Predict the efficiency for a resistance load of 300 kW.
(c) Predict the efficiency for a 300-kVA load at 0.6 pf lagging.
(d) Outline the procedure you would follow to predict the voltage regulation under the condition of part (c).
(e) Carry out the calculations of part (d).

29. A 20-kVA, 4000:200-V transformer has the following parameters: $G = 30\ \mu\mho$, $B = -50\ \mu\mho$, $R = 0.06\ \Omega$, and $X = 0.08\ \Omega$ (Fig. 22.15). Predict the efficiency at full load.

30. For the transformer of Exercise 29:
(a) Draw a large-scale phasor diagram for a unity-pf load ("full load").
(b) Predict the voltage regulation.

31. A 100-kVA distribution transformer has a maximum efficiency of 96% at 70% of rated load. It operates at 70% load 8 hr/day and at 35% load 8 hr/day and at no load 8 hr/day. If electrical energy is 1¢/kWh and annual fixed charges (interest, depreciation, etc.) are 20% of purchase price, about how much extra cost would be justified for a similar 98% efficient transformer?

32. The transformer of Exercise 16 is rated at 5 kVA. Predict the voltage regulation and efficiency under full load at unity pf.

33. A 5000:500-V, 100-kVA, 60-Hz transformer has a no-load input of 2 kW. $R_1 = 5\ \Omega$, $X_1 = 10\ \Omega$, $R_2 = 0.05\ \Omega$, $X_2 = 0.1\ \Omega$.
(a) Draw an appropriate circuit model.
(b) If the transformer delivers 100 A at 500 V to a 0.8 leading power factor load, predict the efficiency.
(c) For a 100-kVA, 0.4 lagging pf load, predict the efficiency.
(d) Predict the voltage regulation for full load at 0.6 lagging pf.
(e) Predict the voltage regulation for full load at 0.6 leading pf.

34. A transformer is labeled: 20 kVA, 120:240 V, 60 Hz. It is assumed to be well designed with optimum use of wire size and core size. (*Note:* Any practical insulation will stand 500 V.) You are asked if it can operate satisfactorily under the following conditions:

	Frequency (Hz)	Primary (V)	Secondary (V)
(a)	60	120	240
(b)	60	60	120
(c)	60	240	480
(d)	120	120	240
(e)	120	240	480
(f)	30	120	240
(g)	30	60	120

PROBLEMS

1. A low-resistance 50-turn coil is wound on a silicon sheet steel core with cross-section of 20 cm^2 and mean length 50 cm. Determine the coil voltage for: (a) A current $i = 1 \cos 400t$ A, and (b) A current $i = 3 + 1 \cos 400t$ A, and compare the results. (c) Sketch a graph of inductive reactance versus dc current for this "saturable core reactor."

2. Using dimensional analysis, derive an expression for K_e in Eq. 22.12 in terms of lamination resistivity and thickness. (Hint: See Examples in Chapter 1. Let P_e be the loss per unit volume. Determine the dimensions of K_e, resistivity ρ, and thickness x. Let $[K_e] = [\rho]^a[x]^b$ and evaluate a and b by inspection of the dimensional equation.)

3. Predict the voltage ratio $\mathbf{V}_2/\mathbf{V}_1$ for the unusual transformer in Fig. 22.24 where $N_2 = N_1$.

4. A coil of N turns is wound on a magnetic core of negligible reluctance. A tap (connection) is made N_2 turns from the end to form an *autotransformer*, and a resistance R is connected across the N_2 turns. A voltage V (rms) is applied across the entire winding of N turns. Sketch the circuit and predict the current I_R in the resistance and the input current I.

5. Design a transformer to operate on regular

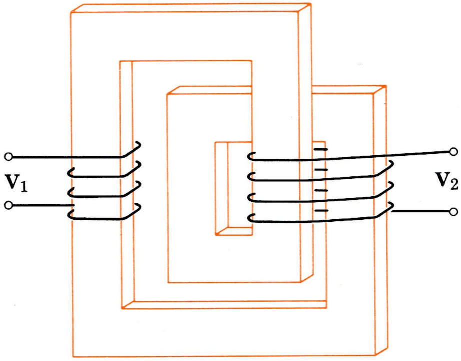


Figure 22.24

house current and supply 2.4 kW to a 60-Ω resistor. Available is a laminated silicon steel core similar to Fig. 22.8a; the cross-section is 30 cm^2 and the mean length is 60 cm. State all assumptions.

(a) Estimate the maximum allowable flux density without an excessive value of H.
(b) Specify the number of turns for the primary.
(c) Estimate the rms value of the no-load primary current.
(d) Specify the number of turns on the secondary.
(e) Predict the full-load primary current.
(f) What size wire should be used? (Consult a handbook.)

6. Show that for symmetrically arranged coils and equal current density in primary and secondary windings $R_2 = a^2R_1$ in Fig. 22.19.

7. Prove that in a transformer maximum efficiency occurs when the variable losses are equal to the fixed losses.

8. For most economic operation, the transformers of Exercise 27 are to reach their maximum efficiency of 98% at three-quarters of full load at unity pf. Specify the critical design parameters (Fig. 22.19) to meet this requirement.

9. Design (i.e., specify the parameters of) a 50-kVA, 4000 : 250-V, 60-Hz transformer to have a full-load efficiency of 98% and a voltage regulation of approximately 2%.

Chapter 23
Principles of Electromechanics

Translational Transducers
Rotational Transducers
Moving-Iron Transducers

In electromechanical energy conversion the coupling may be through the medium of an electric field or a magnetic field. Practical devices employ a variety of physical phenomena. The $q\mathcal{E}$ force on a charge in an electric field is used in a cathode-ray tube for electron acceleration, in a condenser microphone for acoustoelectric conversion, and in proposed space craft for propulsion. The *piezoelectric* effect, whereby certain crystalline materials distort under the action of electric fields, is used in crystal pickups and microphones. The *electrostriction* effect, whereby dielectric materials change dimension in the presence of electric fields, is used in vibration instruments. The analogous deformation of ferromagnetic materials, called *magnetostriction,* is used in supersonic underwater signaling. At the present time, these phenomena are limited in application to rather special situations.

In contrast, the generation of an electromotive force in a moving conductor in a magnetic field and the reverse process of development of a mechanical force by sending a current through a similar conductor find very wide application in instrumentation, automation, and power generation. These electro*magneto*-mechanical phenomena are emphasized in this book, but the techniques developed for their analysis are useful in analyzing the behavior of devices employing other phenomena as well.

The purpose of this chapter is to present basic principles and to show their application to elementary forms of practical devices. Familiar techniques are used to derive linear models for devices employing translation and rotation. New methods are introduced for analyzing the behavior of iron elements moving in magnetic fields. The emphasis throughout is on principles and techniques useful in subsequent chapters for predicting the behavior of practical machines.

Translational Transducers

A device for changing energy from electrical to mechanical form or vice versa is called an electromechanical *transducer.* Usually such a transducer can operate as a *motor* changing electrical energy to mechanical, or as a *generator* changing mechanical energy to electrical. Let us consider how a conductor moving in a magnetic field performs these functions.

Electromotive and Mechanical Forces

In Fig. 23.1 a conductor of length l is moving at velocity u normal to a magnetic field of density B directed into the paper. Each charged particle in the conductor experiences a force

$$\mathbf{f}_q = q(\mathbf{u} \times \mathbf{B}) \qquad (23\text{-}1)$$

The force on a positive charge q is in the direction of advance of a right-hand screw if $\mathbf{u}$ is imagined to rotate into $\mathbf{B}$. By this rule, positive charges are forced upward and negative charges downward, and the net result (whether this is a metal or a semiconductor) is a difference of potential across the conductor. What is the magnitude of the emf generated?

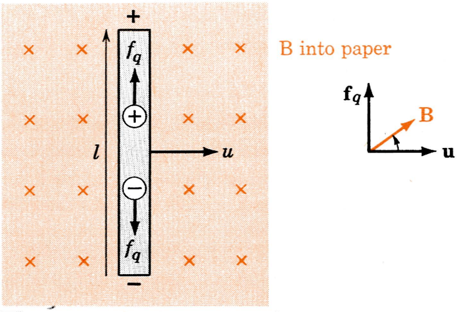


Figure 23.1 Moving-conductor transducer.

The magnitude of the emf e can be determined by noting that charge separation continues until the force of the electric field produced is just equal to the force of the magnetic field causing it. Under equilibrium, the electric force per unit charge (or voltage gradient) is just equal to the magnetic force per unit charge or

$$\frac{\text{Electric force}}{\text{Charge}} = \mathcal{E} = \frac{e}{l} = \frac{\text{magnetic force}}{\text{charge}} = \frac{quB}{q} = Bu \qquad (23\text{-}2)$$

For this case where **u, l,** and **B** are mutually perpendicular, the emf is

$$e = Blu \qquad (23\text{-}3)$$

Since lu is the area "swept" per unit time, one interpretation of Eq. 23-3 is that the emf is just equal to the rate at which magnetic flux is being "cut" by the conductor. In general, the emf, a scalar, is given by the vector expression

$$e = \mathbf{B} \cdot (\mathbf{l} \times \mathbf{u}) = Blu \cos \alpha \sin \beta \qquad (23\text{-}4)$$

where α is the angle from **B** to the normal to the plane containing **l** and **u,** and β is the angle from **l** to **u.** In most engineering devices the elements are arranged in the optimum orientation with $\alpha = 0°$ and $\beta = 90°$, and Eq. 23-3 applies.

If a path is provided for current flow (Fig. 23.2), there is a component of charge velocity u' in the direction of positive current and the charges experience a force f_d that is transmitted to the atoms in the lattice of the conductor and therefore to the conductor itself. Since $qu' = il$ (see Eq. 21-4), the *developed force* is

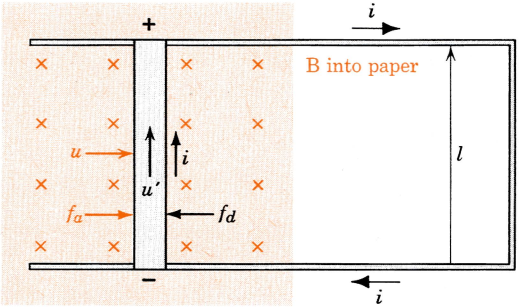


Figure 23.2 A conductor moving on rails.

$$\mathbf{f}_d = i(\mathbf{l} \times \mathbf{B}) \tag{23-5}$$

directed to the left. In general, $f_d = Bli \sin \gamma$ where γ (gamma) is the angle from **l** to **B**. For **l** perpendicular to **B**, the usual case, $\gamma = 90°$ and the magnitude of the developed force is

$$f_d = Bli \tag{23-6}$$

To cause motion in the direction of u, there must be an equal and opposite *applied* force f_a as shown.

Note that the current i produces a magnetic field and this must be considered in some practical machines; here we assume that it is negligibly small compared to B. Note also that the conductor of length l moving at velocity u sweeps out an area $dA = lu\,dt$ in time dt. For uniform B, the rate of change of flux linked by the one-turn closed path in Fig. 23.2 is

$$\frac{d\lambda}{dt} = \frac{d\phi}{dt} = B\frac{dA}{dt} = Blu = e \tag{23-7}$$

As Faraday would have predicted, the emf is just equal to the rate of change of flux linkage.

Bilateral Energy Conversion

With a path provided for current flow the moving conductor is a source of electrical energy, and

$$\text{Electrical power} = P_E = ei = Blu\,i \tag{23-8}$$

The necessary mechanical input is positive since the applied force is in the direction of the velocity, or

$$\text{Mechanical power} = P_M = f_a u = Bli\,u \tag{23-9}$$

As expected from the principle of energy conservation, the mechanical power in newton-meters per second is identically equal to the electrical power in watts.

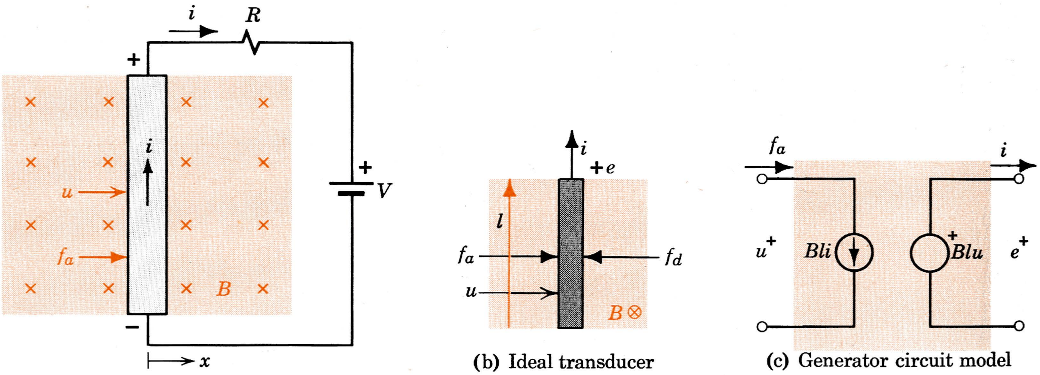
The generated emf can be used to charge a battery as shown in Fig. 23.3. In this case

$$i = \frac{e - V}{R} \tag{23-10}$$

For $e = Blu > V$, current i is positive and the transducer is a *generator.* At a lower velocity, $e < V$, current i is negative, and the developed force is reversed. The conductor is forced to the right, the mechanical power input is negative, and the transducer is acting as a *motor.* The moving conductor transducer is said to be *bilateral;* the energy flow may proceed in either direction.

The Transducer as a Two-Port

Here, just as in the transformer, the magnetic field provides the coupling between input and output ports. Using familiar techniques, we should be able to derive a circuit model for this two-port device. The fact that mechanical force is dependent on a variable in the electrical portion of the system and electromotive force is dependent on a variable in the mechanical portion suggests the use of controlled sources. One possibility is shown in Fig. 23.3c.



(b) Ideal transducer

(c) Generator circuit model

(a) Motor–generator

Figure 23.3 A translational transducer.

This ideal transducer is what Lynch and Truxal[†] call an "electromechanical ideal transformer." In this representation, force is analogous to current and velocity is analogous to voltage (see Table 5-1 and Fig. 22.19b). (What corresponds to the turn ratio?) The polarity of the *force generator Bli* is consistent with the reference direction of i in the *generator* representation shown.

[†] W. A. Lynch and J. G. Truxal, *Principles of Electronic Instrumentation,* McGraw-Hill Book Co., New York, 1962, p. 477.

A real transducer can be represented by considering the mass M of the conductor and the coefficient of sliding friction D. For a *motor*, the governing equation for the mechanical side would be

$$Bli - M\frac{du}{dt} - Du = f_a \tag{23-11}$$

Considering the terminal voltage v, the resistance R, and also the inductance L of the conductor, the governing equation for the electrical side would be

$$v - L\frac{di}{dt} - Ri = e = Blu \tag{23-12}$$

What circuit configuration would satisfy this set of equations?

The symbols shown in Fig. 23.4 and Table 23-1 are conventional for representing M and D in mechanical circuits. Application of d'Alembert's principle (p. 41) to the parallel mechanical circuit yields Eq. 23-11; application of Kirchhoff's voltage law to the series electrical circuit yields Eq. 23-12. We conclude that

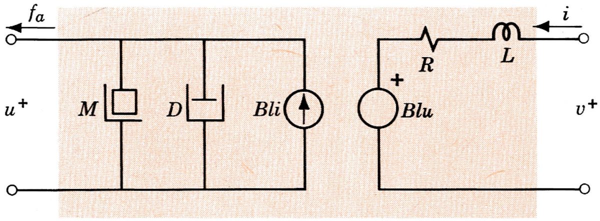
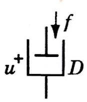
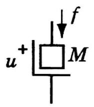
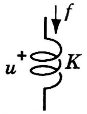
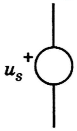
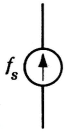


Figure 23.4 Model of a real transducer (motor).

Table 23-1 Conventional Symbols for Mechanical Circuits

Element	Resistance (mechanical)	Mass	Compliance	Velocity source	Force source
Unit	$\frac{\text{Newton-second}}{\text{meter}}$	Kilogram	$\frac{\text{Meter}}{\text{Newton}}$	$\frac{\text{Meter}}{\text{Second}}$	Newton
Symbol					
Characteristic	$f = Du$	$f = M\frac{du}{dt}$	$f = \frac{1}{K}\int u\,dt$	$u = u_s$	$f = f_s$

this circuit model represents the electromechanical system described by the equations. This model is analogous to the hybrid parameter model proposed for a real transformer (Fig. 22.15). (What corresponds to the core loss? To the exciting current?) Note that the ideal transducer represents the *reversible* portion of the energy conversion occurring in a real transducer.

The Dynamic Transducer

The *moving-coil* or *dynamic* transducer is a practical form of the translational transducer. It is used in such devices as loudspeakers, disk recording heads, and phonograph pickups. The essential elements of a dynamic loudspeaker are shown in Fig. 23.5. The permanent magnet establishes an intense radial field in the annular air gap (shown in section). The moving coil consists of a few turns of fine wire wound on a light form and supported in the air gap. The coil form is attached to a light, stiff cone supported by a flexible suspension that keeps the cone and coil in place without restricting the axial motion.

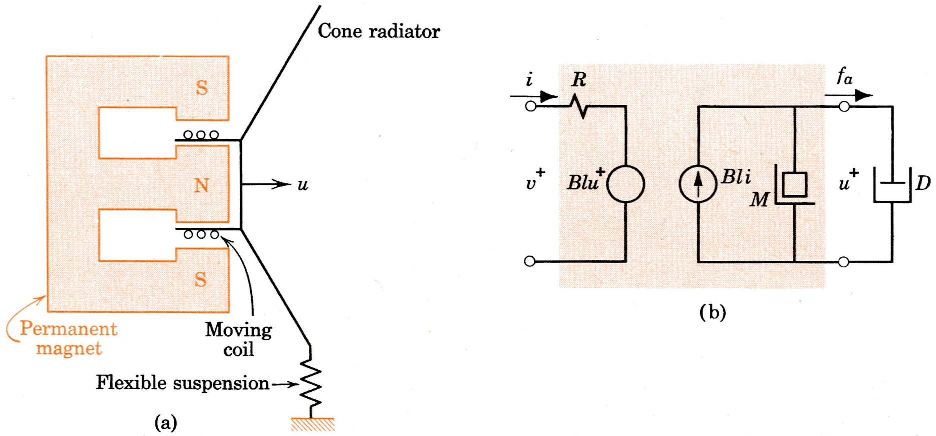


Figure 23.5 Dynamic loudspeaker construction and a simplified circuit model.

An alternating current in the so-called *voice* coil develops a force which is transmitted to the cone. The moving cone produces air pressure waves and acoustic energy is radiated. In the simplified circuit model of Fig. 23.5b, the inductance of the moving coil and the effect of the cone suspension have been neglected. The elements M and D represent inertia effects and the loading effect of the air. At low frequencies the cone moves as a rigid piston, and the electrical energy is efficiently converted into sound waves if the cone is sufficiently large. At higher frequencies the effective mass of the cone presents a large mechanical admittance and the velocity is reduced.† In some high-fidelity systems the low-frequency signals are reproduced by a speaker with a large cone, the *woofer,* and the high-frequency signals are reproduced by the smaller *tweeter.* A *cross-over* network separates the sound into two bands; the signals below 1000 Hz, say, are filtered out by a frequency-selective network and sent to the woofer.

---

† As the frequency increases, the cone radius becomes a larger fraction of a wavelength; therefore, the energy radiated remains high until the cone "breaks up" and no longer moves as a piston.

Example 1

The voice coil of a dynamic loudspeaker consists of 20 turns of diameter 2 cm in a field of $B = 0.5$ T. The effect of air loading on the cone can be represented by a mass M and a resistance coefficient D.

Derive an electric circuit model for determining frequency response. For low and medium frequencies neglect the inductance of the voice coil and the effects of the flexible suspension and work from the simplified circuit model of Fig. 23.5b. To facilitate analysis, refer all mechanical quantities to the electrical side.

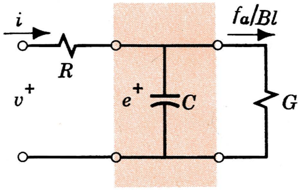


Figure 23.6 An electric circuit model of a dynamic loudspeaker.

Bearing in mind the comparable process in a transformer, we note that the energy stored in the mass M could be represented by an equal energy stored in the analogous capacitance C. Therefore,

$$\tfrac{1}{2}Mu^2 = \tfrac{1}{2}Ce^2 = \tfrac{1}{2}C(Blu)^2$$

Solving,

$$C = \frac{Mu^2}{(Blu)^2} = \frac{M}{(Bl)^2}$$

as shown in Fig. 23.6. The power dissipated in element D is represented by an equal power dissipation in the analogous conductance G. Therefore,

$$f_a u = u^2 D = e^2 G = (Blu)^2 G$$

Solving,

$$G = \frac{u^2 D}{(Blu)^2} = \frac{D}{(Bl)^2}$$

For this device the "turn ratio" Bl is

$$Bl = BN\pi d = 0.5 \times 20 \times \pi \times 0.02 = 0.2\pi \text{ T-m}$$

and a force f_a in the mechanical side corresponds to a current f_a/Bl in the electrical side.

The model indicates a drop in response as the frequency increases.

The Dynamic Pickup

The dynamic loudspeaker mechanism can also be used for a microphone. In fact, in some intercommunication systems the same device serves as microphone and speaker. Sound waves striking the cone cause motion of the coil in the magnetic field. In a well-designed microphone, the emf produced is an accurate replica of the sound signal for frequencies from 40 to 10,000 Hz.

In the dynamic phonograph pickup (Fig. 23.7), the vertical movement of the needle is transmitted to a tiny coil supported in the field of a permanent magnet. The small voltage generated is then amplified in a *preamplifier* that raises the signal above the *hum* level and also corrects the frequency distortion inherent in this type of pickup. The hum is due to voltages induced in the cable from pickup to amplifier by changing magnetic fields set up around transformers and other ac components. The frequency distortion is due to the fact that this type of pickup is *velocity-sensitive,* as illustrated in Example 2.

Example 2

A dynamic phonograph pickup consists of a 20-turn coil (length of each turn = 1 cm) moving normal to a field of B = 0.2 T. If the maximum allowable recording amplitude is 0.02 mm, predict the output voltage at 1000 Hz and at 100 Hz.

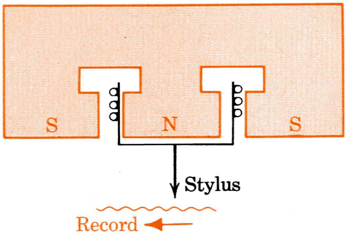


Figure 23.7 A dynamic phonograph pickup for vertical recordings.

The effective length of conductor in the moving coil is $l = 20$ turns $\times$ 1 cm/turn = 20 cm = 0.2 m. For a sinusoidal displacement of amplitude 2×10^{-5} m,

$$x(t) = A \sin \omega t = 2 \times 10^{-5} \sin \omega t$$

and the velocity at 1000 Hz is dx/dt or

$$u = A\omega \cos \omega t = 2 \times 10^{-5} \times 2\pi \times 10^3 \cos \omega t \text{ m/s}$$

Neglecting internal impedance, the output voltage is

$$v = e = Blu = 0.2 \times 0.2 \times 4\pi \times 10^{-2} \cos \omega t \text{ V}$$

The rms value of output voltage at 1000 Hz is

$$V = \frac{0.2 \times 0.2 \times 4\pi \times 10^{-2}}{\sqrt{2}} \cong 0.0035 \text{ V} = 3.5 \text{ mV}$$

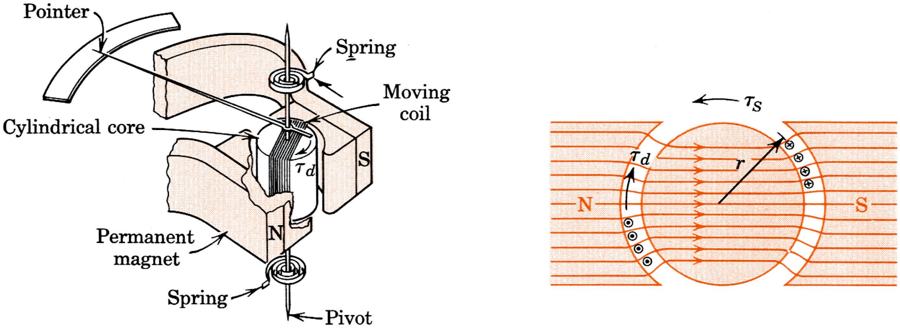
At 100 Hz the output voltage for the same recording amplitude is only $\frac{1}{10}$ as great. Since the amplitude is limited by the distance between grooves, some frequency compensation in the preamplifier is desirable.

Rotational Transducers

In practice, *translational* transducers are limited to applications involving vibratory motion with relatively low displacement. To develop appreciable voltages, velocities must be high and this means relatively high frequencies. In contrast, an electromechanical transducer employing *rotation* of conductors in a magnetic field can produce large low-frequency or direct voltages (or torques) in easily constructed configurations.

The d'Arsonval Mechanism

The mechanism of the ordinary dc ammeter was invented nearly a hundred years ago, but it is still one of our best current-measuring devices. The d'Arsonval movement is widely used in instrumentation, and it is a good illustration of a simple rotational transducer.



(a) Construction (b) Cross section

Figure 23.8 The d'Arsonval meter movement.

As shown in Fig. 23.8, a moving coil is wound on a rectangular form that is supported on jeweled bearings in the air gap between the poles of a permanent magnet and a stationary iron core. Current flowing in the coil produces forces that tend to rotate the coil assembly. The current is conducted to and from the coil by two fine springs that also supply a restoring torque to oppose rotation of the coil. A lightweight pointer attached to the assembly indicates the angular deflection on a calibrated scale.

Analysis of the behavior of the mechanism is simplified if we note that rotation is restricted, by means of mechanical stops, to the region of uniform radial magnetic field. If there are N turns ($2N$ conductors) of length l normal to the magnetic field at a radius r, the torque τ_d (tau) developed by a current i is

$$\tau_d = f_d \cdot r = B(2Nl)i \cdot r = 2NBlri = k_M i \tag{23-13}$$

The combined effect of the two spiral springs is a restoring torque τ_S which is proportional to angular deflection or

$$\tau_S = \frac{\theta}{K_R} \tag{23-14}$$

where K_R is the rotational compliance in radians (or degrees) per newton-meter. Under equilibrium conditions, the restoring torque is just equal to the developed torque, and

$$\theta = 2NBlrK_R \cdot i = k_\theta i \tag{23-15}$$

The deflection is directly proportional to the current in the moving coil, and the meter scale is linear.

Example 3

A d'Arsonval ammeter (Fig. 23.8) has a rectangular coil form with length $l = 1.25$ cm and width $w = 2r = 2$ cm. Each spiral spring has a compliance $K_R = 16 \times 10^7$ °/N-m. A uniform flux density of 0.4 T is supplied by the permanent magnet. Determine the number of turns so that full-scale deflection of 100° is obtained with a current of 1 mA.

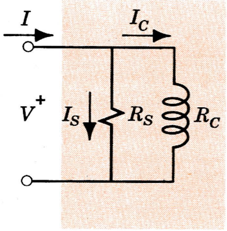
Noting that two springs are only half as compliant as one and then solving Eq. 23-15,

$$N = \frac{\theta}{2Blri(K_R/2)}$$

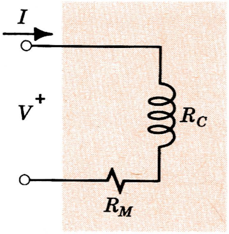
$$= \frac{100}{2 \times 0.4 \times 0.0125 \times 0.01 \times 0.001 \times 8 \times 10^7}$$

$$= 12.5 \text{ turns}$$

(Is it easy to construct a coil with an extra half turn?)



(a) Ammeter



(b) Voltmeter

Figure 23.9 DC meter connections.

Meter Application and Response

The movement in Example 3 provides full-scale deflection with a current of 1 mA. How could it be used in a 0- to 5-A ammeter? In the connection of Fig. 23.9a, some of the meter current is *shunted* through R_S and only a part of the current to be measured flows through the moving coil. If the coil resistance R_C is 20 Ω, then full-scale deflection requires a voltage of

$$V = I_C R_C = 0.001 \times 20 = 0.020 \text{ V}$$

For full-scale deflection with a meter current of 5 A, the current in the *shunt* R_S is $5 - 0.001 = 4.999$ A. The resistance of the shunt should be

$$R_S = \frac{V}{I_S} = \frac{V}{I - I_C} = \frac{0.02}{4.999} \cong 0.004\ \Omega \qquad (23\text{-}16)$$

A sturdy and precisely adjusted resistance may be mounted inside the meter case or be available for external connection.

Could this same movement be used in a voltmeter? The d'Arsonval mechanism is essentially a current-measuring instrument, but it can be used as a voltmeter by adding a series resistance R_M as in Fig. 23.9b. For a 0- to 300-V voltmeter, full-scale deflection is obtained with 1 mA in the moving coil and a voltage drop of 0.02 V; the remainder of the voltage must appear across the *multiplier* resistance. Therefore,

$$R_M = \frac{V_M}{I_C} = \frac{V - V_C}{I_C} = \frac{300 - 0.02}{0.001} \cong 300{,}000\ \Omega \qquad (23\text{-}17)$$

In a *multimeter*, several shunts and multipliers are mounted within the meter case and are connected at will by a selector switch (see Fig. 27.2).

In addition to providing a convenient method of solution, a model should give insight into the general behavior of a device. Let us see what we can learn from the general model of a d'Arsonval movement. Neglecting the small inductance of the moving coil and the friction in the jeweled bearings, the model is as shown in Fig. 23.10a. For rotation, the appropriate mechanical variables are torque τ and angular velocity ω. The element J represents the rotational inertia of the coil assembly and K_R is the combined spring compliance. The mechanical proportionality constant $k_M = \tau_d/i$ is defined by Eq. 23-13. The electrical proportionality constant k_E is defined as

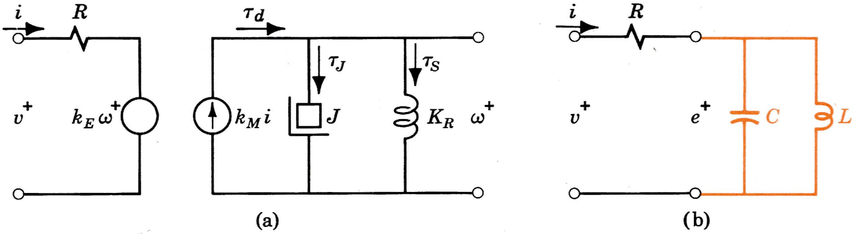


Figure 23.10 Circuit model of a d'Arsonval movement.

$$k_E = \frac{e}{\omega} = \frac{2NBlu}{\omega} = 2NBlr = k_M \tag{23-18}$$

since $u = r\omega$; we see that in this electromechanical transducer also there is a simple "turn ratio."

Using the turn-ratio concept (see Example 1), the rotational admittances on the mechanical side can be referred to the electrical side as in Fig. 23.10b. Suddenly connecting the instrument in a line corresponds to the application of a step current to the circuit model. What is the dynamic response of the movement? Using the approach of Chapter 4, we can express the admittance of the *parallel portion* as

$$Y(s) = sC + \frac{1}{sL} = \frac{s^2LC + 1}{sL} \tag{23-19}$$

Because the admittance function indicates purely imaginary zeros, we expect an undamped sinusoidal response.

If the only mechanical resistance is that due to bearing friction, the needle will indeed oscillate widely. How could the long wait for a steady reading be reduced? There are two commonly used methods for introducing the necessary damping. One is to mount a large but light vane on the coil assembly and introduce air-friction damping. Another is to make the coil form of a conducting material, such as aluminum. The coil form then becomes a shorted turn, and coil motion induces an emf that results in energy dissipation. (How would this effect be represented in the circuit model of Fig. 23.10? How is the admittance function altered? See Exercise 23.10.)

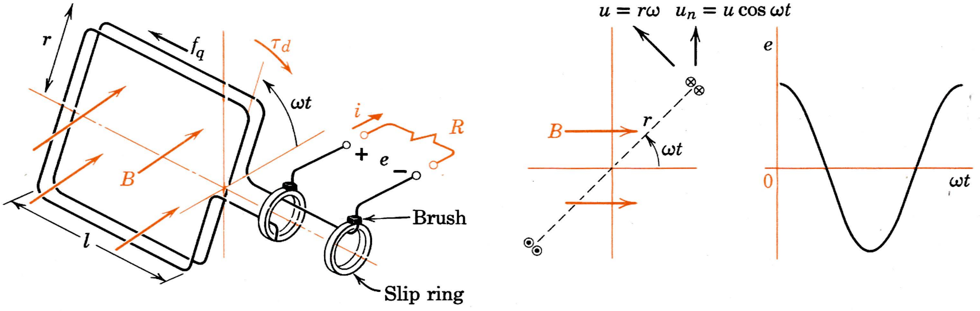
In a properly designed meter movement with optimum damping, the needle quickly reaches a steady reading. In the steady state, $\omega = 0$, $e = k_E\omega = 0$, and $\tau_J = J\,d\omega/dt = 0$. The model then indicates that

$$k_M i = \tau_S = \frac{1}{K_R}\int \omega\, dt = \frac{\theta}{K_R} \tag{23-20}$$

which agrees with Eq. 23-15. The dynamic response of electromechanical transducers is treated in more detail in Chapter 26.

The Elementary Dynamo

Rotation of the coil in a d'Arsonval mechanism is limited to something less than 180°. If we are to realize the advantages of this configuration, we must permit continuous rotation. One possibility is to mount a rectangular coil on a shaft and connect the ends of the coil to conducting *slip rings* attached to the shaft but insulated from it as in the elementary *dynamo* of Fig. 23.11. As the N-turn coil rotates at angular velocity ω in a uniform magnetic field B, the generated emf e is connected to the external circuit by fixed *brushes* sliding on the rotating slip rings.



(a) (b) (c)

Figure 23.11 Elementary dynamo construction and operation.

Electromotive Force. The polarity of the generated emf is determined by considering the force on positive charges in the moving conductors. The right-hand rule indicates that the force f_q on positive charges in the upper conductors in Fig. 23.11b is into the paper; therefore, the front slip ring is negative at the instant shown. (What is the direction of the force in the lower conductors?) The magnitude of the emf can be determined by any of the three expressions employed so far. For $2N$ conductors of length l moving at velocity u in a magnetic field of density B, the emf (Eq. 23-4) is

$$e = 2NBlu \cos \alpha \sin \beta$$

Where $\beta = 90°$ and $2rl$ = area A,

$$e = 2NBlr\omega \cos \omega t = NBA\omega \cos \omega t \tag{23-21}$$

The rate at which flux is being cut by $2N$ conductors (Eq. 23-3) is

$$e = 2NBlu_n = 2NBlr\omega \cos \omega t = NBA\omega \cos \omega t \tag{23-22}$$

The rate of change of flux linkage (Eq. 23-7) is

$$e = \frac{d\lambda}{dt} = NB\frac{d(A \sin \omega t)}{dt} = NBA\omega \cos \omega t \tag{23-23}$$

Equation 23-23 shows that, in a uniform magnetic field, the emf is determined by the area A of the coil and not the particular shape or dimensions. The same emf or torque is developed by a circular coil, say, of the same area. Equations 23-13 and 18 were derived for rectangular coils in a uniform field; a more general formulation is

$$k_M = \frac{\tau_d}{i} = k_E = \frac{e}{\omega} = NBA \tag{23-24}$$

Torque. In the d'Arsonval movement there is a uniform radial field and Eq. 23-24 indicates that

$$e = k_E\omega = NBA\omega \tag{23-25}$$

Comparing Eqs. 23-21 and 24 and reasoning by analogy with Eq. 23-13, we anticipate that in the elementary dynamo

$$\tau_d = k_M i \cos \omega t = NBAi \cos \omega t \tag{23-26}$$

The direction of this torque depends on the direction of the current. If the coil is rotating in the direction shown in Fig. 23.11a and if a resistance R is connected across the terminals, the current will flow as shown and the right-hand rule indicates a clockwise torque. This is a *developed* torque that must be overcome by the mechanical power source. If an electrical power source is used, however, this torque is available for doing mechanical work. Example 4, on the following page, demonstrates that the ideal elementary dynamo is another bilateral electromechanical energy converter.

Example 4

The elementary dynamo of Fig. 23.11 is driven at constant angular velocity by a mechanical source. The output terminals are connected to a load resistance R. Determine the electrical power supplied and the mechanical power required.

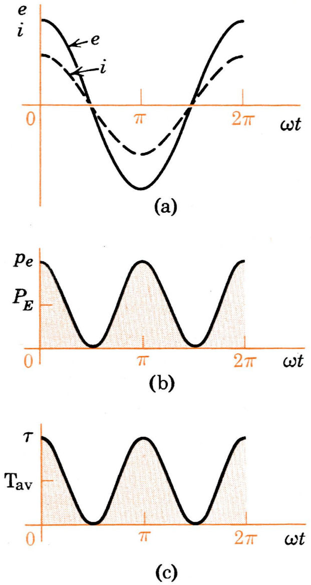


Figure 23.12 Power and torque relations in an elementary dynamo.

Assuming that the resistance and inductive reactance of the coil are small compared to R, all the emf appears across R and the current is

$$i = \frac{e}{R} = \frac{NBA\omega \cos \omega t}{R} = I_m \cos \omega t$$

The average electrical power is

$$P_E = \left(\frac{I_m}{\sqrt{2}}\right)^2 R = \frac{(NBA\omega)^2}{2R}$$

Assuming that the friction torque is negligibly small, the applied torque must be equal and opposite to the developed torque; its magnitude (see Fig. 23.12) is

$$\tau_a = NBA \cos \omega t \cdot i = NBA \cos \omega t \frac{NBA\omega \cos \omega t}{R}$$

The average mechanical power is the product of the average torque times the angular velocity of rotation or

$$P_M = \mathrm{T}_{\mathrm{av}}\omega = \frac{1}{2}\frac{\omega(NBA)^2}{R}\omega = \frac{(NBA\omega)^2}{2R} = P_E$$

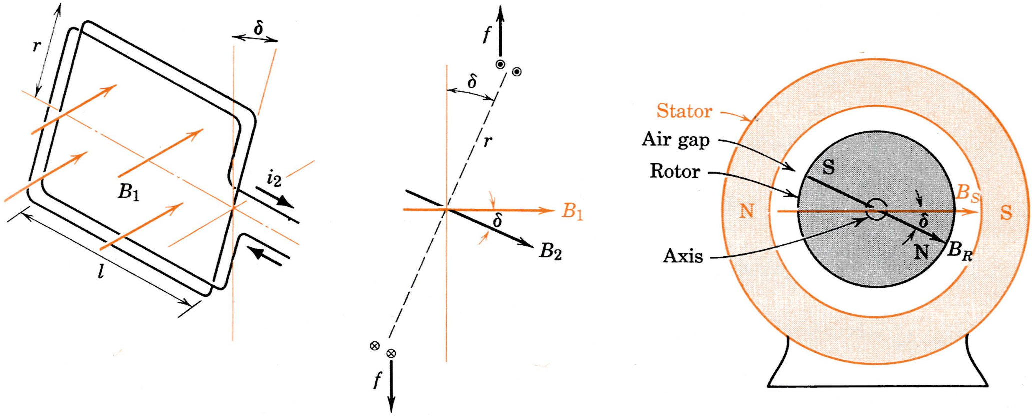
Under these idealizing assumptions, all the mechanical power is converted to electrical power.

A General Torque Equation

As preparation for the analysis of a variety of rotating machines we need a general expression for torque. A single, simple relation that would apply to ac or dc machines employing electro- or permanent magnets in various configurations would be very convenient. Let us take another look at the elementary dynamo and try to abstract the essential idea.

As shown in Fig. 23.13b, a current i_2 in the upper conductor of length l in a field of uniform magnetic flux density B_1 develops a force

$$\mathbf{f} = i_2(\mathbf{l} \times \mathbf{B}_1)$$



(a) (b) (c)

Figure 23.13 Torque on a coil in a magnetic field.

and the torque developed by an N-turn coil of width $2r$ is

$$\tau_d = 2NB_1li_2 \cdot r\sin\delta = B_1A \cdot Ni_2 \cdot \sin\delta$$
$$= k_aB_1i_2\sin\delta \tag{23-27a}$$

where δ (delta) is the angle between $\mathbf{B}_1$ and the normal to the plane of the coil. But what is Ni_2? This is the mmf of the coil and establishes a flux density $\mathbf{B}_2$ in the direction shown. An alternative form of the torque expression is

$$\tau_d = k_bB_1B_2\sin\delta \tag{23-27b}$$

In words, the torque can be considered as due to the *interaction of an electric current and a magnetic field* or to the *interaction of two fields*. But B_1 itself is just a manifestation of a current i_1, either in a coil or in the atoms of a magnet, so the torque can also be considered as due to the *interaction of two currents*. A third formulation of the basic relation is

$$\tau_d = k_ci_1i_2\sin\delta \tag{23-27c}$$

The angle δ is called the *torque angle* or *power angle*. The torque increases with $\sin\delta$ and reaches a maximum at $\delta = 90°$. Figure 23.13c symbolizes the general rotating machine. The stationary portion or *stator* sets up the field B_S and the rotating portion or *rotor* sets up the field B_R. Either or both of these fields may be established by electric currents in a practical machine. For a given current (or quantity of copper) and a given flux (or quantity of iron), the machine designer realizes the maximum torque if he arranges for δ to be 90°. For continuous torque, sin δ must not change sign. These are critical considerations and dictate the mechanical arrangement and electrical connection of most practical rotating machines.

Rotating Machines

A rotating electromechanical converter has several important advantages. Torques and velocities can be constant instead of pulsating, and high velocities are possible without the high accelerating and decelerating forces inherent in translational devices. High velocities permit high voltages and, therefore, higher electrical power per unit weight. Because frictional losses increase rapidly with velocity, the bearings are placed on shafts of small radius (and low linear velocity) and the active conductors are at a much larger radius.

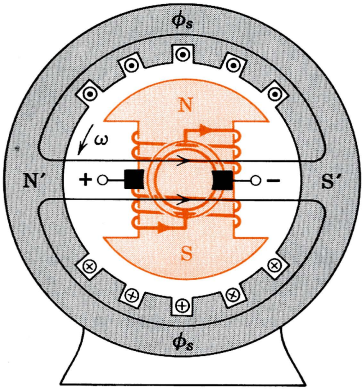
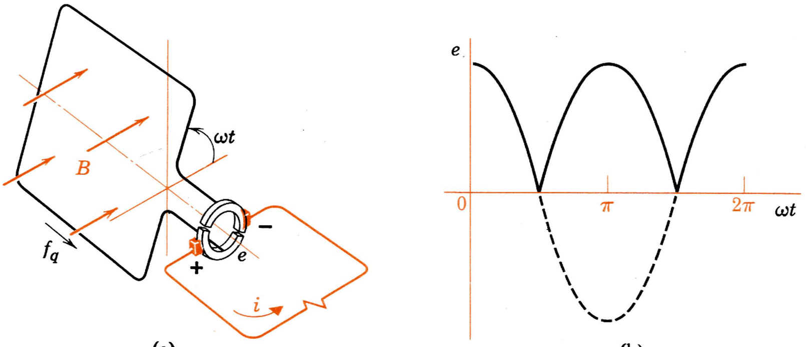


Figure 23.14 An alternator.

AC Machines. A sinusoidal emf is generated when the coil of an elementary dynamo is turned at constant angular velocity in a uniform magnetic field. A more effective arrangement of copper and iron is shown in Fig. 23.14. Direct current is supplied to the field windings on the rotor through brushes and slip rings. The small air gap minimizes the magnetizing current required for a given flux density. The ac emf ($e = d\lambda/dt$) is generated in conductors connected in series and placed in slots in the stator. Having the high-power conductors stationary minimizes the problems of insulation, centrifugal force, and electrical connection. By distributing the turns along the periphery of the stator and shaping the air gap properly, the output voltage can be made to approximate a sinusoid. Most of the world's electrical power is generated in more elaborate versions of this basic alternator.

Under what conditions can the alternator function as a motor? If the field is energized to establish B_R and a direct current is supplied to the stator to establish B_S, the rotor will turn until $\delta = 0$ and then stop. When the fields are aligned, sin $\delta = 0$ and there is no torque; this is a position of equilibrium. But motoring action is possible if we can upset the equilibrium in just the right way. Assume that at a particular instant the current is out of the upper conductors and into the lower conductors in Fig. 23.14. This creates a stator field with poles N′ and S′ as shown and $\delta = 90°$. The developed torque tends to turn the rotor clockwise (in the $-\omega$ direction). Because of inertia the rotor will rotate past the position of $\delta = 0$. If at that instant the current in the stator is reversed, the new torque is also clockwise and the rotor continues to turn. The stator current can be reversed by a mechanical switch, or an alternating current of just the right frequency can be used. The latter arrangement is the basis of the *synchronous motor* treated in more detail in Chapter 25.

DC Machines. To obtain unidirectional torque from direct currents, we use a rotating mechanical switch called a *split-ring commutator.* The same device permits the generation of direct currents in a rotating machine. In dc machines, the *field* winding is stationary and the assembly of conductors and commutators, called the *armature,* rotates. The operation of the commutator is indicated in Fig. 23.15. At the instant shown, the coil emf is decreasing. At $\omega t = \pi/2$, the emf is zero and the coil is shorted by the brushes. An instant later the lower conductor begins to move up across the flux and the generated emf is reversed; but the lower conductor has automatically been switched to the negative brush and the terminal polarity is unchanged. The commutator could be called "a synchronous full-wave rectifier."

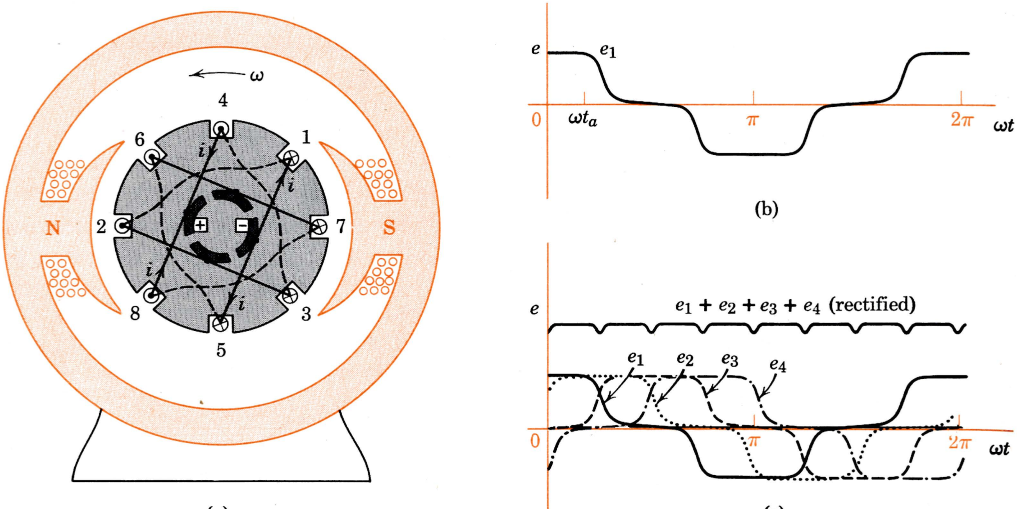


(a) (b)

Figure 23.15 Commutator operation and generated voltage waveform.

In the more practical form of Fig. 23.16a, the armature conductors are embedded in a cylindrical iron rotor. The stationary field windings establish a high and nearly uniform flux density in the small air gap. The emf generated in a single conductor (number 1) is shown in Fig. 23.16b; instant ωt_a corresponds to the left-hand figure with conductor 1 just leaving the S pole. Starting at the negative brush (drawn on the inside of the commutator but actually riding on the outside) and going around the circuit, we see that conductors 1, 2, 3, and 4 are connected in series. The sum of the individual emfs appears at the positive brush. Going in the other direction from the negative brush, we see that the other four conductors are connected in series. In this winding there are two *parallel paths.* As the armature rotates, carrying the segmented commutator with it, conductor 1 is switched out of one path and conductor 8 takes its place.

In a practical machine there are many conductors and many commutator



(a) (b) (c)

Figure 23.16 Two-pole dc generator construction and voltage waveforms.

segments. At any instant many coils are connected in series between brushes and a few coils are short-circuited (by the brushes) as they are being switched from one circuit to another. As each coil is switched, there is a momentary drop in terminal voltage, resulting in what is called commutator *ripple,* but this is small compared to the total effect of many coils in series.

As indicated in Fig. 23.16a, the current is into the paper in all conductors moving upward, and out of the paper in all conductors moving downward. The net result of all the currents is a flux density distribution that can be represented by a vector directed upward. As the armature rotates, this flux density *remains fixed in space* at the optimum torque angle $\delta = 90°$. The resulting developed torque is clockwise and, in a generator, must be overcome by the mechanical power source. If direct current from an electrical source is introduced at the brushes, a constant unidirectional torque is developed and the machine functions as a motor.

The two-pole commutator machine resembles the d'Arsonval mechanism in many ways. As would be expected from Eq. 23-18, for constant flux density in the air gap

$$K_M = \frac{\mathrm{T}}{I} = K_E = \frac{E}{\Omega} \tag{23-28}$$

where the capital letters indicate direct or average values of torque T, armature current I, emf E, and angular velocity Ω. If total flux per pole Φ is a variable (by controlling field current and, thereby, flux density), the expressions for torque and emf are

$$\mathrm{T} = K\Phi I \quad \text{and} \quad E = K\Phi\Omega \tag{23-29}$$

The steady-state model of a commutator machine is shown in Fig. 23.17. In the steady state, energy storage terms are not significant and only the rotational mechanical friction D_R and electrical resistance R are considered. The value of K depends on the machine construction and the electrical connection, as illustrated in Example 5.

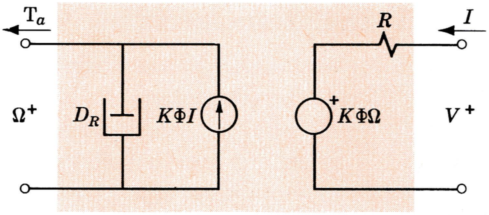


Figure 23.17 Steady-state model of a commutator machine (motor).

Example 5

In a dc machine (Fig. 23.18), the armature is wound on a laminated iron cylinder 15 cm long and 15 cm in diameter. The N and S pole faces are 15 cm long (into the paper) and 10 cm along the circumference. The average flux density in the air gap under the pole faces is 1 T. If there are 80 conductors in series between the brushes and the machine is turning at $n = 1500$ rpm, predict the no-load terminal voltage.

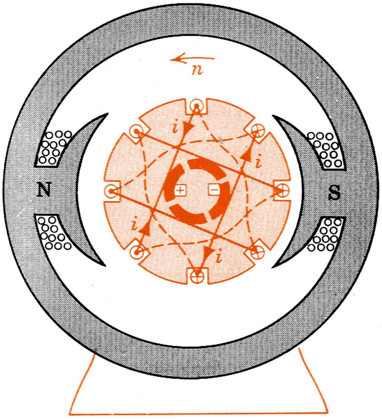


Figure 23.18 Voltage calculation.

The dc voltage can be predicted in any of the three ways expressed in Eqs. 23-21, 22, and 23. In this case the rate of flux cutting is a convenient approach. For N_s conductors (*not* coils) in series, the average emf is

$$E = N_s \frac{\Delta\Phi}{\Delta t} \frac{\text{Wb}}{\text{s}} = N_s \Phi \frac{\text{Wb}}{\text{pole}} p \frac{\text{poles}}{\text{rev}} \frac{n}{60} \frac{\text{rev}}{\text{sec}}$$

$$= \frac{N_s \Phi p n}{60} \text{ volts} \qquad (23\text{-}30)$$

Here the flux per pole is

$$\Phi = BA = 1 \times 0.15 \times 0.1 = 0.015 \text{ Wb}$$

Therefore,

$$E = 80 \times 0.015 \times 2 \times \frac{1500}{60} = 60 \text{ V}$$

Equation 23-30 can be derived directly from Eq. 23-22 by taking into account the number of poles and recognizing that the dc voltage is just the average of a rectified sinusoid. The value of K in Eq. 23-29 can be derived from Eq. 23-30 by noting that $\Omega = 2\pi n/60$. Solving,

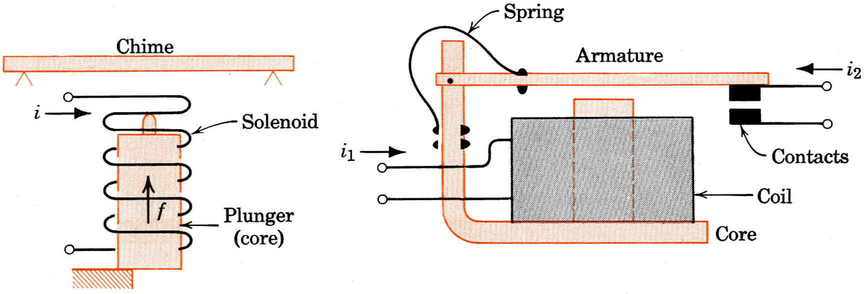
$$K = \frac{E}{\Phi\Omega} = \frac{N_s p}{2\pi} \qquad (23\text{-}31)$$

The number of conductors in series is conveniently determined from $N_s = Z/p'$ where Z is the total number of conductors and p' is the number of parallel paths. In one common type of winding, $p' = p$.

Moving-Iron Transducers

Most of the devices described so far employ a moving conductor in a magnetic field. The elementary alternator of Fig. 23.14 differs in that the iron rotor, which could be a permanent magnet, moves past the stationary conductors; but this is essentially just a convenient method of obtaining relative motion of the conductors with respect to the field. Another type of electromagnetomechanical energy converter employs an initially unmagnetized iron member moving in the direction of a magnetic field. Such *moving-iron* transducers cannot be analyzed conveniently in terms of forces on currents; a different approach is required.

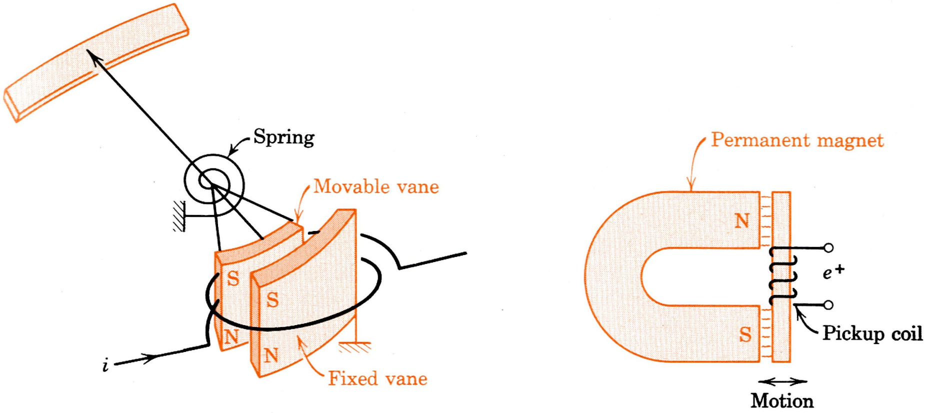
In the simple door chime of Fig. 23.19a, part of an iron core or plunger is within the solenoid. When a current i flows in the solenoid, the plunger is accelerated upward and strikes the chime; to design such a chime we need to know the force on the core as a function of position. A relay is a control device whereby the current in circuit 2 can be turned on or off by a current in circuit 1. In Fig. 23.19b, a current i_1 energizes the coil and moves the armature against the restraining force of the spring. The movable contact meets the fixed contact and circuit 2 is closed. Mechanical stops (not shown) limit the armature travel. Relay specifications include the *pickup* or closing current and the *dropout* or opening current. We shall learn how these currents can be predicted.



(a) Door chime (b) Relay

Figure 23.19 Common examples of moving-iron transducers.

The *iron-vane* mechanism of Fig. 23.20a is another form of the moving-iron transducer. Current in a multiturn coil magnetizes two soft-iron elements. Since the polarities are the same and like poles repel, the movable vane rotates with respect to the fixed vane. The developed torque, opposed by a spring, is always in the same direction; therefore, this is a good mechanism for measuring alternating currents. In a *reluctance pickup* (Fig. 23.20b), a vibratory motion of the small iron armature changes the air gap and, therefore, the reluctance of the magnetic circuit. The flux established by the permanent magnet changes and a voltage is induced in the pickup coil. We need to know whether the generated emf is proportional to the displacement, velocity, or acceleration of the armature.



(a) Iron–vane instrument **(b) Reluctance pickup**

Figure 23.20 Other moving-iron transducers.

Virtual Work Method

One approach to the quantitative analysis of an electromechanical transducer employs the principle of *virtual work.* We are interested in determining a force; therefore, we assume that a small displacement takes place, calculate the work that would be done in such a *virtual* displacement, and then determine the actual force.

If the plunger of Fig. 23.19a were to move an infinitesimal vertical distance dx, the work done on the plunger would be $f\,dx$. Where does this energy come from? It must come from the electrical input. Does all the electrical input appear as mechanical work? No, some of it appears as heat from electrical resistance and mechanical friction, and some of it may be stored in the magnetic field. Let us minimize i^2R by using high-conductivity wire and minimize mechanical friction by proper design. Then practically all the electrical input appears as either mechanical work or as an increase in energy stored in the magnetic field.

In a linear magnetic system, the energy stored is $\frac{1}{2}Li^2$; if the current is constant at a value I, the change in energy stored is $\frac{1}{2}I^2\,dL$. (Does the inductance change as the core moves? Yes, because the reluctance changes.) The electrical input is

$$p\,dt = ei\,dt = (N\,d\Phi/dt)I\,dt = I\,d(N\Phi)$$

where $d(N\Phi)$ is the change in flux linkage associated with a virtual displacement dx. But $N\Phi = LI$ (Eq. 22-6) and if I is constant, $I\,d(N\Phi) = I^2\,dL$. Expressing the idea in the previous paragraph, we write

$$\text{Electrical input} = \text{work done} + \text{increase in energy stored} \qquad (23\text{-}32)$$

or

$$I^2\,dL = f\,dx + \tfrac{1}{2}I^2\,dL$$

Solving,

$$f = \frac{I^2\,dL - \frac{1}{2}I^2\,dL}{dx} = \frac{1}{2}I^2\frac{dL}{dx} \tag{23-33}$$

This is a relation of great simplicity and, it turns out, of considerable generality.

> *The force is directly proportional to the rate of change of inductance and in the direction associated with an increase in inductance (since I^2 is always positive).*

Reasoning by analogy, what result would you anticipate for the torque in a rotational transducer? A similar analysis yields

$$\tau = \frac{1}{2}I^2\frac{dL}{d\theta} \tag{23-34}$$

Another basic conclusion is that in such a linear magnetic system, half the electrical energy input is converted to mechanical energy and the other half appears as an increase in the energy stored in the magnetic field.

Example 6

A magnetic circuit is completed through a soft-iron rotor as shown in Fig. 23.21. Derive an expression for torque as a function of angular position.

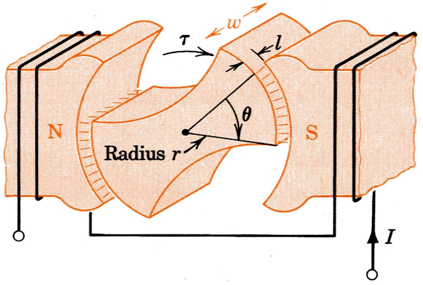


Figure 23.21 Reluctance torque.

To simplify the analysis, let us assume:

1. All the reluctance of the magnetic circuit is in the air gaps of length l.
2. There is no fringing so the effective area of each gap is the area of overlap $= A = r\theta w$.

The total reluctance for two gaps in series is

$$\mathcal{R} = \frac{2l}{\mu_o A} = \frac{2l}{\mu_o r w \theta}$$

For an N-turn coil, the inductance is

$$L = \frac{N\Phi}{I} = \frac{N\mathcal{F}}{I\mathcal{R}} = \frac{N^2 I}{I\mathcal{R}} = \frac{N^2\mu_o r w\theta}{2l}$$

By Eq. 23-34 the torque is

$$\tau = \frac{1}{2}I^2\frac{dL}{d\theta} = \frac{\mu_o N^2 I^2 r w}{4l}$$

or the torque is independent of θ under the assumed conditions. Actually, there is some positive torque for negative values of θ. (Why?) Also note that there is no torque when overlap is complete because $dL/d\theta$ drops to zero.

The *reluctance torque* evaluated in Example 6 is important in practical machines. Small electric-clock motors operate on this principle. Some medium-size synchronous motors are pulled into synchronism by reluctance torque, and in large machines the reluctance effect contributes a significant part of the total torque.

Energy Balance Method

Equation 23-32 is an abbreviated *energy balance* based on the law of conservation of energy. In general, the energy added to a system goes to increase the energy stored within the system or appears as an output from the system (Fig. 23.22). Some of the chemical energy of the fuel supplied to a diesel-electric system is stored in the kinetic energy of the flywheel, or as thermal energy evidenced by an increase in temperature, and the remainder appears as useful electrical output or unavailable heat radiated or carried away in the exhaust. In dealing with electromagnetomechanical converters, it is convenient to write the energy balance as

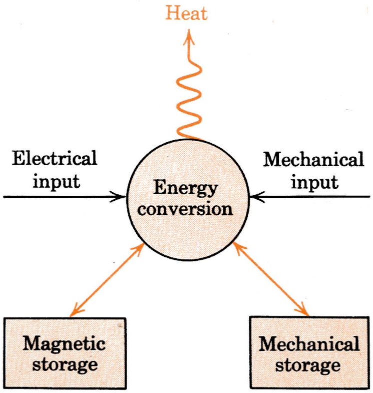


Figure 23.22 An energy conversion system.

$$\begin{matrix}\text{Mechanical}\\ \text{input}\end{matrix} + \begin{matrix}\text{electrical}\\ \text{input}\end{matrix} = \begin{matrix}\text{mechanical}\\ \text{storage}\end{matrix} + \begin{matrix}\text{magnetic}\\ \text{storage}\end{matrix} + \text{heat} \qquad (23\text{-}35)$$

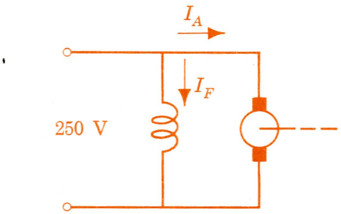
This equation applies to all types of converters. In a generator, mechanical input is positive and electrical input is negative. In a d'Arsonval movement, electrical input is positive, mechanical input is zero, and mechanical energy is stored in the springs. During current changes, the kinetic energy of the moving coil may be significant. In a relay, the change in energy stored in the air gap just balances the mechanical output (see Eq. 23-36). Also, in every real device some energy is converted into heat in irreversible processes. Copper loss, hysteresis and eddy-current losses, and bearing and air-friction losses must all be considered. The energy-balance approach is particularly useful in situations in which overall effects are more easily handled than detailed phenomena.†

---

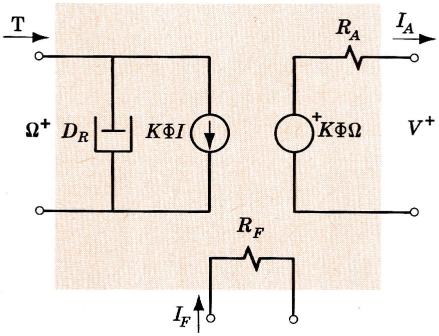
† See Chapters 4 and 5 of H. H. Skilling, *Electromechanics*, John Wiley & Sons, New York, 1962.

Example 7

A commutator machine is rated at 5 kW, 250 V, 2000 rpm. The armature resistance R_A is 1 Ω. Driven from the electrical end at 2000 rpm, the no-load power input to the armature is $I_A = 1.2$ A at 250 V with the field winding ($R_F = 250$ Ω) excited by $I_F = 1$ A. Estimate the efficiency of this machine as a 5-kW generator.



(a) Wiring diagram



(b) Steady-state model

Figure 23.23 Generator efficiency.

In Fig. 23.23, input power $I_F^2R_F = 1^2 \times 250 = 250$ W is required to provide the necessary magnetic flux. This is power lost and it appears as heat.

In no-load steady-state operation there is no output and no change in energy storage; therefore, the armature input of 1.2 A at 250 V = 300 W is all loss. A small part ($I_A^2R_A = 1.2^2 \times 1 = 1.44$ W) is copper loss, but most of the input power goes to supply air, bearing, and brush friction and the eddy-current and hysteresis losses associated with flux changes in the rotating armature core. All these losses are dependent on speed, but are nearly independent of load.

At full-load of 5 kW, $I_A = 5000\text{ W}/250\text{ V} = 20$ A and the armature copper loss is

$$I_A^2R_A = 20^2 \times 1 = 400\text{ W}.$$

With the generator driven from the mechanical end, the rotational losses can be assumed to be equal to the no-load armature input and are represented by D_R in the model. Then the energy balance (Eq. 23.35) is

$$\begin{matrix}\text{Mech.}\\ \text{input}\end{matrix} + \begin{bmatrix}\text{field} & - & \text{elec.}\\ \text{input} & & \text{output}\end{bmatrix}$$

$$= 0 + 0 + \begin{bmatrix}\text{field} & + & \text{arm.} & + & \text{rotat.}\\ \text{loss} & & \text{loss} & & \text{loss}\end{bmatrix}$$

$$\text{Mech. input} + [250 - 5000] = [250 + 400 + 300]$$

Solving,

$$\text{Mech. input} = 5000 + 400 + 300 = 5700\text{ W}$$

(Field input is all loss and appears on both sides.)

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{\text{elect. output}}{\text{mech. input} + \text{elect. input}}$$

$$= \frac{5000}{5700 + 250} = 0.84 \text{ or } 84\%$$

The Electromagnet

A versatile example of the moving-iron transducer is the electromagnet shown in Fig. 23.24. An electric current, alternating or direct, develops a pull on the pivoted armature and raises the mass M a distance dl. Motion of the armature may be used to close or open contacts, to open or close valves, or to engage mechanisms. The important design criterion is the force developed for a given current.

Considering the shaded region as an electromechanical system, we see that there is electrical energy input and mechanical energy output, we suspect that

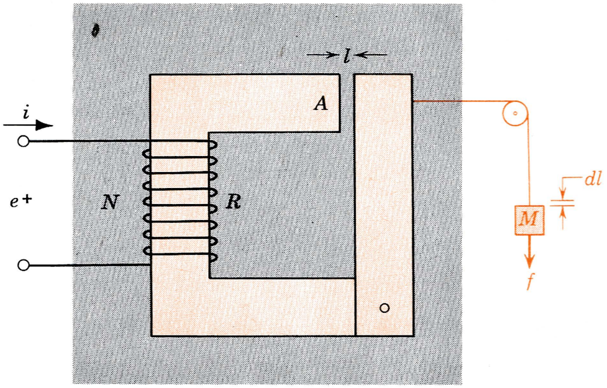


Figure 23.24 An electromagnet with a pivoted armature.

there is energy stored, and we know there is energy lost irreversibly. In terms of the energy-balance equation, the mechanical input for a virtual displacement dl is $-(f\,dl)$ since the work is actually done *by* the system. If during this infinitesimal displacement we adjust the current so that the magnetic flux remains constant, $N\,d\phi/dt$ is zero and there is no induced emf in the coil. Then the electrical input is just $i^2R\,dt$, where dt is the time required for displacement dl. (How are these conditions different from those imposed in deriving Eq. 23-33?)

If the motion of the armature is horizontal, there is no change in potential energy stored; and if the motion is slow, there is no change in kinetic energy stored. The energy stored in the magnetic field does change. From Eq. 22-9 we know that the energy per unit volume is $\frac{1}{2}B^2/\mu$. If ϕ and B are constant, there is no change in the energy stored in the iron; but the energy stored in the air gap *decreases* by $(\frac{1}{2}B^2/\mu_o)A\,dl$, where A is the area of the gap if fringing is neglected.

The final term in the energy balance is primarily the $i^2R\,dt$ loss in the coil winding. With a frictionless pivot there is no mechanical friction loss, and with no change in ϕ there is no eddy-current or hysteresis loss. (Was this a convenient condition?) Equation 23-35 becomes

$$-f\,dl + i^2R\,dt = 0 - \frac{1}{2}\frac{B^2}{\mu_o}A\,dl + i^2R\,dt \tag{23-36}$$

Actually, current i is a variable, but since the same term appears on both sides of the equation its evaluation is unimportant. Solving for the developed force in newtons,

$$f = \frac{B^2A}{2\mu_o} \tag{23-37}$$

This result is general and does not require that ϕ (and B) be constant (see Exercise 27). Since the flux density possible in practical materials is limited by saturation, this relation places an upper limit on the force developed by an electromagnet of a given cross section.

To determine the relation between force and current, we assume that the permeability of the iron is so high that all the mmf in the magnetic circuit appears across the reluctance of the air gap. Then

$$B = \frac{\phi}{A} = \frac{\mathcal{F}}{\mathcal{R}_a A} = \frac{Ni}{(l/\mu_o A)A} = \frac{\mu_o Ni}{l}$$

and Eq. 23-37 becomes

$$f = \frac{\mu_o N^2 i^2 A}{2l^2} \tag{23-38}$$

where μ_o = the permeability of free space = $4\pi \times 10^{-7}$ H/m,
N = the number of turns on the coil,
i = the current in amperes,
A = the area of the air gap in meters^2, and
l = the length of the gap in meters.

Note that the force increases as the armature moves and the air gap length decreases. For very small air gaps the assumptions must be reexamined.

Example 8

In the relay shown in Fig. 23.25, the contacts are held open by a spring exerting a force of 0.2 N. The gap length is 4 mm when the contacts are open and 1 mm when closed. The coil of 5000 turns is wound on a core 1 cm^2 in cross-section. Predict the pickup and dropout currents for this relay.

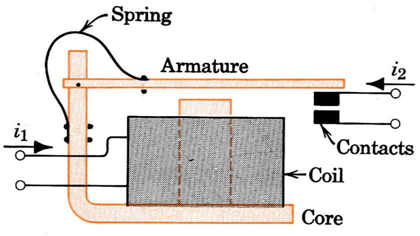


Figure 23.25 Relay performance.

To simplify the solution, we assume:

1. All reluctance is in a uniform air gap.
2. Fringing is negligible so that $A = 10^{-4}$ m^2.
3. Spring force is constant and acts at a distance equal to half the distance from the pivot to the air gap.
4. Friction and inertia effects are negligible.

For pickup or closing, the gap length is 4×10^{-3} m and the required force is $\frac{1}{2} \times 0.2\text{ N} = 0.1$ N. Solving Eq. 23-38 for current,

$$i = \sqrt{\frac{2l^2 f}{\mu_o N^2 A}} = \sqrt{\frac{2 \times 16 \times 10^{-6} \times 0.1}{4\pi \times 10^{-7} \times 25 \times 10^6 \times 10^{-4}}}$$
$$\cong 0.032\text{ A or }32\text{ mA}$$

Since current is directly proportional to gap length, the opening or dropout current is approximately $32/4 = 8$ mA.

To check the adequacy of the core area and the validity of the reluctance assumption, we note that

$$B = \frac{\mu_o Ni}{l} = \frac{4\pi \times 10^{-7} \times 5 \times 10^3 \times 32 \times 10^{-3}}{4 \times 10^{-3}}$$
$$= 0.05\text{ T a very conservative value.}$$

SUMMARY

- A conductor moving in a magnetic field

  generates an emf $e = \mathbf{B} \cdot (\mathbf{l} \times \mathbf{u}) = Blu \cos \alpha \sin \beta$

  and develops a force $\mathbf{f} = i(\mathbf{l} \times \mathbf{B})$ where $f = Bli \sin \gamma$

  The translational transducer is a bilateral energy converter that can be represented by a two-port model with

  $$P_E = ei = P_M = fu = Bliu$$

  The dynamic transducer used in loudspeakers and pickups consists of a coil in translation in a transverse magnetic field.

- The d'Arsonval mechanism widely used in instrumentation consists of a coil in rotation in a radial magnetic field; the coil generates an emf $e = NBA\omega$ and develops a torque $\tau = NBAi$.

- A coil rotating at constant speed in a uniform magnetic field generates an emf $e = NBA\omega \cos \omega t$ and develops a torque $\tau = NBAi \cos \omega t$.
  The emf can be evaluated in terms of the velocity of moving conductors, the rate at which flux is cut, or the rate of change of flux linkage.
  The torque can be evaluated in terms of the interaction of: two electric currents, two magnetic fields, or a current and a field.

- A useful form of the general torque equation is $\tau = kB_R B_S \sin \delta$.

- A practical alternator, with direct current supplied to the rotor, generates a sinusoidal emf in the stator.
  The alternator can operate as a motor if the rotor turns at synchronous speed and alternating current is supplied to the stator.

- The commutator permits the generation of direct current and the production of unidirectional torque; as the rotor turns, conductors are automatically switched to the proper circuit.

  $$\frac{\mathrm{T}}{I} = \frac{E}{\Omega} = K\Phi = \frac{N_s p}{2\pi}\phi$$

- The method of virtual work and the energy-balance concept are particularly useful in analyzing moving-iron transducers; in general,

  $$f = \tfrac{1}{2}I^2 \frac{dL}{dx} \quad \text{and} \quad \tau = \tfrac{1}{2}I^2 \frac{dL}{d\theta}$$

- The force developed by an electromagnet is approximately

  $$f = \frac{B^2 A}{2\mu_o} = \frac{\mu_o N^2 i^2 A}{2l^2}$$

REVIEW QUESTIONS

1. List four basically different reversible electromechanical phenomena.
2. On a transducer, terminal A is positive with respect to terminal B. If an ammeter indicates current is flowing *out* of terminal B, is this transducer acting as a motor or as a generator? Explain.
3. Explain how, in a moving current-carrying conductor in a magnetic field, positive charges are forced in one direction and the conductor in another.
4. Outline the steps in deriving an expression for emf generated in a moving conductor.
5. In Fig. 23.4, what corresponds to the turn ratio in a transformer?
6. Draw and label a mechanical circuit model to represent a device governed by the equation $Bli = M\,du/dt + Du + \int u\,dt/K$.
7. An advertisement says: "The superior tone quality is due to the highly compliant suspension and massive magnet structure of the speaker." Analyze this statement. What does it mean? Are you impressed?
8. A loudspeaker has two coils; one consists of a few turns in an air gap and the other consists of many turns on an iron core. What function is served by each?
9. What frequency-response characteristic is desirable in the preamplifier following a dynamic phonograph pickup?
10. Why is the current to a d'Arsonval instrument carried by the spiral springs? What is the maximum possible deflection?
11. Write an expression for sensitivity (°/A) for a d'Arsonval mechanism.
12. How can the same instrument measure voltage and current?
13. Explain the concept of a "turn ratio" in a rotating coil transducer.
14. How is the circuit of Fig. 23.10b altered when the movement is damped?
15. Write three different expressions for the emf generated in a coil moving with respect to a permanent magnet.
16. Write an expression for torque in an ideal dynamo that does not include any magnetic factors. Explain the energy basis.
17. What is the torque angle? How does it affect machine design?
18. What advantages do rotational devices have over translational devices?
19. What condition is necessary if a machine with direct current in the rotor and alternating current in the stator is to develop useful torque?
20. Draw a sketch of an actual commutator (automobile generator or starter motor) showing segments, insulation, and brushes.
21. Explain the operation of a commutator in a dc motor.
22. Are there eddy-current or hysteresis losses in a dc generator? Why?
23. Given a bar of soft iron in a magnetic field; is the torque in such a direction as to increase or decrease the stored magnetic energy?
24. What is the meaning of "virtual"? What is "virtual work?"
25. Explain the operation of a reluctance clockmotor.
26. Make an energy balance for a dc generator that is accelerating; clearly identify each term.
27. How would the energy balance (Eq. 23-36) differ for current held constant?
28. How is the lifting force of an electromagnet affected by a slight reduction in pole-face area?

EXERCISES

1. Assume that a magnetic field of density $B = 0.1$ T is emerging vertically from your desk surface. Holding a pencil 15 cm in length inclined at an angle of 30° from the desk, you move your hand laterally at a speed of 60 cm/s to the right. Predict the magnitude and polarity of the voltage generated in the pencil lead.
2. In Fig. 23.26, a long, straight conductor carries a current I in air.

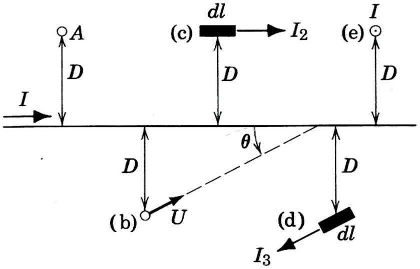


Figure 23.26

   (a) Express the magnetic flux density and the magnetic field intensity at point A and indicate their directions.
   (b) Express the force on an electron b moving with velocity U as shown and indicate its direction.
   (c) Repeat for element c of length dl carrying current I_2.
   (d) Repeat for element d.
   (e) Repeat for a similar element e carrying current into the paper.
3. Derive Eq. 23-2 by equating the work done *by* an external agency to the work done *on* the moving charges.
4. In a simplified form of magnetohydrodynamic generator (see Fig. 20.6) the electrodes are parallel and of length $L = 30$ cm. A field of $B = 1$ T (into the paper) exists in the channel which has a square cross-section $b = 10$ cm on a side. The plasma has a resistivity $\rho = 10^{-5}$ Ω-m and an average velocity $U = 10$ m/s.
   (a) Explain why an emf should be developed across the electrodes.
   (b) Derive a literal expression for the electrode voltage indicated by a high-resistance voltmeter.
   (c) Estimate the maximum power obtainable from this generator under the given conditions.
   (d) What nonelectrical measurements would reveal this power?
5. In Fig. 23.27 a conducting bar moves with constant speed u (so that its position is $x = ut$) in a region of linearly increasing magnetic flux density into the paper. A resistance R is connected between the high conductivity rails. Determine:
   (a) $i(t)$, the current in the bar.
   (b) $f(t)$, the force on the bar.

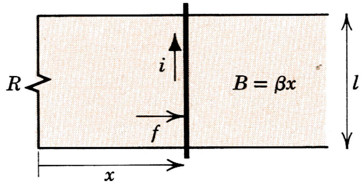


Figure 23.27

6. In the transducer of Fig. 23.2, the resistance of the rails is negligible compared to the resistance R of the conductor of mass M which slides with friction coefficient D on the rails. The conductor is initially at rest and a constant force f_a is applied at time $t = 0$.
   (a) Draw an appropriate circuit model.
   (b) Determine the steady-state velocity U.
7. An electromechanical two-port (Fig. 23.28) consists of a bar of resistance R and mass M restrained by a spring of compliance K

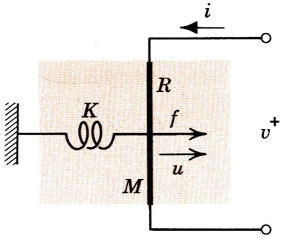


Figure 23.28

in a field of flux density B. Write the governing equations for the electrical and mechanical sides, and devise an appropriate circuit model.

8. In a dynamic loudspeaker similar to Fig. 23.5a, a magnetic flux density of 0.6 T is available across the annular air gap of radius 1 cm, length 2 mm, and depth 1 cm (in the direction of u). A maximum force of 0.8 oz is to be developed by a current $i = 0.2 \cos \omega t$ A. Estimate the required mmf to provide the design flux density, and specify the number of turns on the voice coil.

9. A dynamic phonograph pickup is similar to the speaker in Fig. 23.5. The lateral displacement of the needle is transmitted directly to the 20-turn coil that vibrates in a field of $B = 0.1$ T. Each turn of the coil is 3 cm long.
   (a) How many grooves per centimeter are there on a typical 33-rpm record? (What is the playing time?)
   (b) If approximately 80% of the space per groove is available for recording, what is the maximum peak-to-peak displacement allowable?
   (c) What is the maximum voltage (rms value) generated in this pickup at 100 Hz? At 10,000 Hz?
   (d) For a uniform response (output voltage versus frequency), what provisions in regard to amplitude of vibration should be made in recording?

10. An ammeter similar to Fig. 23.8 is to deflect "full-scale" (2 rad) with a current of 1 mA in the square coil, 2 cm on a side. A flux density of 0.4 T is available from a permanent magnet, and the compliance of each spring is 10^5 rad/N-m. Complete the design by specifying the remaining parameter.

11. A d'Arsonval movement provides full-scale deflection with a current of 5 mA or an applied voltage of 30 mV. Specify the shunt and multiplier to provide ranges of 0 to 1 A and 0 to 50 V.

12. Repeat Exercise 11 for ranges of 0 to 5 A and 0 to 300 V.

13. The circuit model of an undamped meter movement is shown in Fig. 23.10a. One method of introducing damping is to make the coil form of a conducting material such as aluminum.
   (a) How does this introduce the desired damping?
   (b) Redraw Fig. 23.10a to include the effect of the coil form.
   (c) Redraw Fig. 23.10b and derive the expression (similar to Eq. 23-19) for admittance of the parallel portion of the circuit.
   (d) Define the condition for critical damping of the mechanical portion of the movement.

14. An elementary dynamo (Fig. 23.11a) consists of an N-turn coil of length l and radius r meters turning at n rpm in a field of B teslas. A resistance R is connected across the output.
   (a) Express the generated voltage $e(t)$.
   (b) For an applied torque $\tau_a = \mathrm{T}_m \cos \omega t$, determine the output current and the mechanical power input.

15. A dc generator similar to Fig. 23.16, but with four stator poles, turns at 2000 rpm. Each pole face (at the air gap) is 10 cm × 12 cm and $B = 1$ T. Between brushes there are 75 turns effectively in series, and each turn has a total length of 42 cm.
   (a) Estimate the emf generated.
   (b) For an output current of 50 A, estimate the torque developed and the mechanical power input.

16. In the dc generator of Exercise 15, the effective air gap is 3 mm and there are 500 turns per field pole (see Fig. 21.9b). Stating any necessary assumptions, estimate the values of k_a, k_b, and k_c in Equation 23-27.

17. (a) If continuous torque is to be developed in the machine in Fig. 23.16a, what can you say about the field established by the rotor?
   (b) The machine of Fig. 23.14 is to be used as a motor. DC current in the rotor will establish a magnetic field defined by N and S rotating at speed ω. For continuous torque, what can you say about the magnetic field N′–S′ in the stator?
   (c) The stator of the machine in Fig. 23.14 provides a moving field rotating at

speed ω. If the rotor is replaced by the rotor in Fig. 23.16a, what can you say about the field in the rotor if continuous torque is to be developed?

18. An alternator similar to Fig. 23.14 but with 6 poles on the rotor is driven at 1200 rpm.
   (a) Sketch the alternator, labeling the poles, and predict the output frequency.
   (b) With no output current, the required torque is 5 N-m; with an output current of 20 A, the required torque is 45 N-m. Neglecting stator resistance, estimate the generated emf.

19. The alternator of Fig. 23.14 generates an emf $e = 120\sqrt{2} \cos 2\pi 60t$ V.
   (a) Calculate the speed in rpm.
   (b) Predict the new emf if the speed is increased by 10%.
   (c) At the original speed, predict the emf if the current in the rotor winding is decreased by 10%.
   (d) If 10 additional slots are cut in the unslotted portion of the stator and another identical winding installed, predict the emf $e_2(t)$ generated in the new winding under the original conditions.

20. In Fig. 23.16b, the emf generated at 1200 rpm by a single conductor has an average value of 4 V.
   (a) Estimate the total flux per pole and the maximum emf.
   (b) Predict the dc output voltage (no load) if each path consists of 60 conductors in series.
   (c) If a load current of 10 A flows, predict the electrical power generated and the counter torque developed.

21. The field winding of a 10 kW, 6-pole, 220-V dc generator produces 10 mWb of magnetic flux per pole. There are 72 armature slots and two conductors are placed in each slot. If the coils are connected internally so that there are two parallel paths (i.e., half the conductors are connected in series), estimate the speed of rotation for rated voltage at no load.

22. A commutator machine with an effective armature resistance of 1 Ω requires an input current of 2 A at 200 V just to turn the rotor at rated speed of 2000 rpm (no load). Draw an appropriate circuit model and predict the mechanical power and torque output for an input of 20 A at 200 V.

23. A magnetically soft iron bar (low residual magnetism) is mounted on an axis (into the paper) between the poles of a magnet (Fig. 23.29).

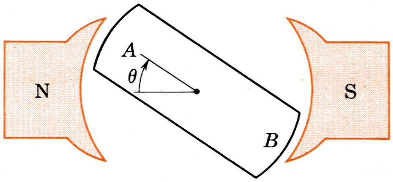


Figure 23.29

   (a) As the bar is rotated through 360°, how does the magnetic polarity of the bar change?
   (b) At what positions is the torque zero? Maximum?
   (c) What position does the bar take if left to swing freely?
   (d) At what position is the total energy stored in the system a minimum?

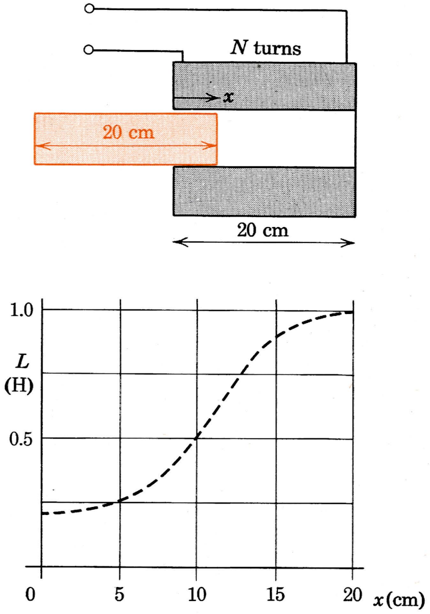


Figure 23.30

24. In the ac ammeter of Fig. 3.22, the total inductance with all coils aligned is 10 mH and with the rotating coil L_R reversed, 6 mH. Assuming that $L = L_S + L_R + 2M \cos \theta$ (where $\theta = 0°$ when the coils are aligned), estimate the torque developed by a current of 2 A for $\theta = -90°$.

25. Figure 23.30 shows the variation of inductance for a solenoid and plunger.
  (a) Specify the required current for a maximum pull of 20 N.
  (b) Plot a curve of pull vs. plunger position.

26. In the reluctance pickup of Fig. 23.20b, the permanent magnet provides a total mmf $\mathfrak{F}$ across the two air gaps of area A and length l. With the N-turn coil open circuited and the armature moving at velocity u, derive expressions for:
  (a) The induced emf e.
  (b) The necessary applied force f_a.

27. In an electromagnet (Fig. 23.24), the current i is held constant. Stating any necessary assumptions, write the energy-balance equation for a small motion dl, and derive Eq. 23-38.

28. (a) Starting with Eq. 23-37 and carefully explaining your reasoning as you go along, derive Eq. 23-33.
  (b) Check Eq. 23-38 dimensionally.

29. A magnetic contactor used to close circuits carrying heavy currents is similar to Fig. 23.24 with contacts mounted on the pivoted armature. The cross-section of the core is 8 cm^2 and the length is 25 cm. Specify the current required in a 200-turn coil to develop a force of 20 N across a 1-cm gap.

30. With a current of 5 A in the actuating coil of the contactor of Exercise 29, estimate the force required to open the contacts when the air gap is zero.

31. The dimensions of the electromagnet in Fig. 23.31 are in cm; the depth of the silicon steel core and armature is 5 cm.

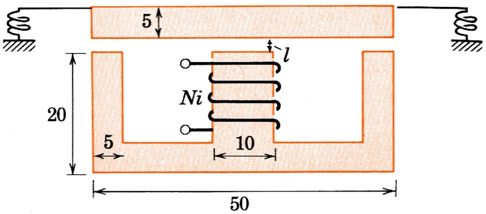


Figure 23.31

  (a) Design a coil (specify N and i) to be placed on the center leg to provide a total pull on the armature (supported by springs) of 60 N at an air gap length $l = 1$ cm.
  (b) For the design values of N and i, estimate the force when the armature is in contact with the core.
  (c) Using the results of parts (a) and (b) and your knowledge of magnetic circuits, sketch a graph of force f as a function of gap length l.

PROBLEMS

1. Small quantities of material are to be moved back and forth inside a hermetically sealed chamber, and the following scheme is proposed. The material of mass M will be on a light trolley running on rails separated distance L; wires attached to the rails will be brought out through seals and connected to a battery of voltage V through a reversing switch and rheostat R. It is expected that the magnetic field of the earth (vertical component $= 3 \times 10^{-5}$ T) reacting with the current in the trolley axles will move the trolley.
  (a) Stating all assumptions, derive expressions for the trolley acceleration, starting from rest, and the time to move a distance D.
  (b) Discuss the feasibility of the proposal.

2. The device sketched in Fig. 23.32 is proposed to pump a hot, corrosive liquid (such as molten sodium in a nuclear reactor).
  (a) What electrical or magnetic properties must the liquid possess?
  (b) For positive current I, how would the device work?
  (c) For $B = 1$ T across the 10-cm × 30-cm

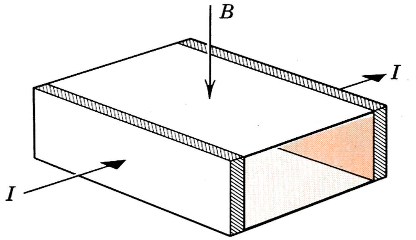


Figure 23.32

pipe, what current would be required for a pressure rise of 5 N/cm^2 (approximately 6 psi)?

3. A d'Arsonval instrument cannot be employed to measure rapidly changing currents. Investigate the response of such an instrument to alternating voltages of the form $v = V_m \cos \omega t$. Work in terms of phasors, using θ for angular displacement and $\dot{\theta}$ for angular velocity to distinguish from frequency ω. Neglect the inductance of the moving coil and the frictional resistance of the bearings.
   (a) Draw an appropriate circuit model, clearly labeled, and derive an expression for the input impedance.
   (b) Sketch the variation in impedance magnitude Z as a function of frequency ω, labeling significant values of Z and ω.
   (c) Derive an expression for the ratio $\theta/\mathbf{V}$ and sketch a graph of the ratio θ_m/V_m as a function of ω.
4. The movement of a 50-mV (full-scale) d'Arsonval meter is characterized by: $R = 50\ \Omega$, $J = 10^{-8}$ N-m-s^2/rad, and K_R (each spring) $= 2 \times 10^6$ rad/N-m.
   (a) Determine the frequency of the undamped oscillation and specify an appropriate value for D_R for reasonable damping.
   (b) Determine and sketch the response $\theta(t)$ of the damped movement to a suddenly applied voltage $v = 25$ mV.
5. You are to derive an expression for the average torque developed by the motor of Fig. 23.33. The reluctance of the magnetic circuit is a minimum at $\theta = 0°$. Assume that the rotor shape is such that $\mathcal{R} = \mathcal{R}_0 - \mathcal{R}_a \cos 2\theta$. (Does this look reasonable?) If the impedance of the winding is negligibly small, $v = e$ and a sinusoidal voltage produces a sinusoidal magnetic flux $\phi = \Phi_m \cos \omega t$. Since unidirectional torque is produced only if the speed of rotation is synchronized with the alternation of the flux, let $\theta = \omega t - \delta$. What is θ when the flux is maximum? If the instantaneous torque is represented by the sum of components of various frequencies, which component corresponds to average torque?

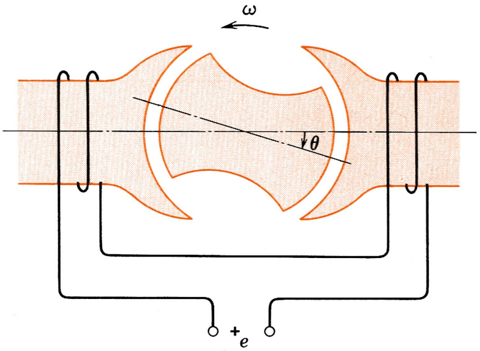


Figure 23.33

6. Prepare a chart summarizing the characteristics of some of the devices studied so far. Rule a page into 4 columns headed Device, Input, Output, and Function. Provide 10 rows for the following devices: Amplifier, Rectifier, Modulator, Transformer, Loudspeaker, Pickup, Ammeter, Relay, Commutator Motor, and Alternator. Complete the chart by filling in the blanks with brief statements (10 words or less).
7. The capacitance of a parallel plate capacitor is $C = \epsilon A/x$ where ϵ = permittivity, A = area (of one plate), and x = spacing.
   (a) Using the method of virtual work (Hint: hold V constant), derive an expression for the force between the plates as a function of voltage V, capacitance C, and spacing x.
   (b) Eliminate C from the expression for force derived in part (a) and interpret physically the sign of the result.
   (c) Repeat part (a) holding charge Q constant and allowing x to change slowly.
8. In the magnetic clutch of Fig. 23.34, the dimensions are: $a = b = d = 1$ cm, $c = 3$ cm. The depth of the silicon steel core is 2 cm and the clutch plate is circular.

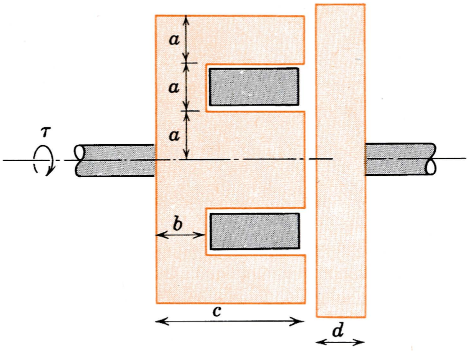


Figure 23.34

(a) For an effective air gap of 0.2 mm and a friction coefficient of 0.9, estimate the torque that can be transmitted.

(b) Estimate the horsepower that can be transmitted at 1000 rpm.

(c) Estimate the required number of ampere turns on the actuating coil.

9. Design an electromagnet for lifting compressed automobiles in a junk yard.

(a) Select a suitable material and a shape that will provide two lifting surfaces with rectangular blocks of compressed metal.

(b) Assume a maximum acceleration of 16.1 ft/s^2 and estimate the necessary lifting force in newtons.

(c) Allowing for paint and surface irregularities, specify the pole face area required and the necessary ampere-turns.

Chapter 24

Direct-Current Machines

Generators
Motors

The rotating machine is preeminent as a means for converting great amounts of electromechanical energy. The design of efficient and economical machines is a highly developed art, demanding specialized training and extensive experience; the number of engineers who *design* electrical machines is small. In contrast, engineers in all branches of engineering and in nearly all functional classifications *use* electrical machines. This chapter and the next are concerned with the use of existing machines in steady-state applications under normal conditions of operation.

Our purpose here is to provide the background necessary to select machines to meet general requirements and to predict the behavior of the machines selected. Electrical machines vary widely in operating characteristics, power requirements, flexibility, ruggedness, and cost; therefore, the features of only the most common types of machines are described in detail. Prediction of the precise behavior of a given machine is quite difficult. By making simplifying assumptions, however, we can derive relatively simple models that give satisfactory results under most conditions. Being aware of the effect of the assumptions made, we can (with more training) modify the results if greater precision is required.

From Chapter 23 we know the principles of emf generation and torque development that apply to all rotating electromagnetic devices. Now we are going to investigate the construction features and operating characteristics of the most important types of machines. Our procedure, as with transistors and transformers, is to derive circuit models and use these to predict steady-state performance. Later (in Chapter 26) we shall study the dynamic behavior of machines as preparation for applications in control systems.

Classification of Electrical Machines

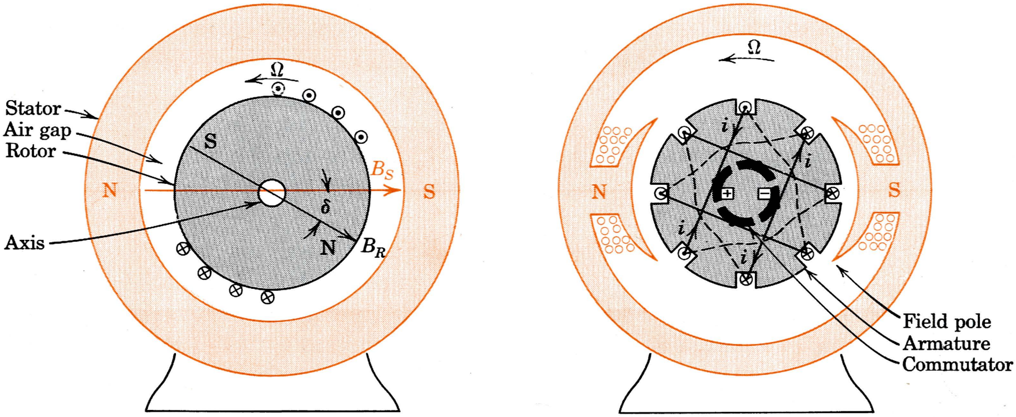
The rotating machine consists essentially of axial conductors moving in a magnetic field established in a cylindrical gap between two iron cores, but in practice it takes on a great variety of forms. One basis of classification is in terms of the current flowing in the stator and rotor windings. On this basis:

Direct-current machines—direct current in both stator and rotor.
Synchronous machines—alternating current in one, direct in the other.
Induction machines—alternating current in both stator and rotor.

The construction of an induction motor is relatively simple, but the prediction of performance is quite complicated and a sophisticated model is required. In contrast, the construction of a dc motor is complicated but a simple model is quite satisfactory. Therefore, we start our study with dc machines.

Direct-Current Generators

The stator or *field* of a dc generator or motor (Fig. 24.1) consists of an even number of magnetic poles (alternating N and S around the circumference) excited by direct current flowing in the field windings. The rotor or *armature* consists of a cylindrical iron core carrying the active conductors embedded in slots and connected to the segments of the *commutator.* Direct current is carried to and from the armature by stationary *brushes* (drawn on the inside of the commutator but actually riding on the outside). The commutator automatically switches the conductors so that the external current from a generator or the torque from a motor is steady and unidirectional. (See Fig. 23.16 and the accompanying explanation.) The location of the brushes ensures that the torque angle δ is 90°.



(a) General machine (b) DC generator

Figure 24.1 Configurations of rotating machines.

Basic Relations

The emf generated in each conductor can be calculated from $e = Blu$ and the total emf is determined by the number of conductors in series at any time.

The torque developed in any conductor can be calculated from $\tau_d = Blir$ and the total torque is the summation of the individual contributions. For a given machine in the steady state, the basic relations are:

$$E = K\Phi\Omega \quad \text{and} \quad \mathrm{T}_d = K\Phi I_A \tag{24-1}$$

where E = the generated emf in volts,
Φ = the air gap flux per pole in webers,
Ω = the angular velocity in radians per second,
T_d = the developed torque in newton-meters,
I_A = the armature current in amperes, and
K = a constant for the given machine (see Eq. 23-31).

Equations 24-1 apply *at the air gap;* the terminal voltage differs from the generated emf by the armature resistance voltage drop and the shaft torque differs from the developed torque by the mechanical resistance torque.

The dc machine is a bilateral energy converter so that at the air gap the developed mechanical power is just equal to the generated electrical power or

$$\mathrm{T}_d\Omega = EI_A \tag{24-2}$$

The air gap power represents only the reversible portion of the electromechanical energy conversion. Losses due to irreversible energy transformations occur in all practical machines.

Circuit Model of a Generator

For a dc generator in the steady state, the circuit model of Fig. 24.2 is applicable. Under steady-state conditions, currents and velocities are constant (or changing very slowly) so that there are no changes in mechanical or magnetic energy storage to consider. This means that the rotational inertia and electrical inductance terms are omitted from the model. The field resistance R_F must be considered in efficiency calculations.

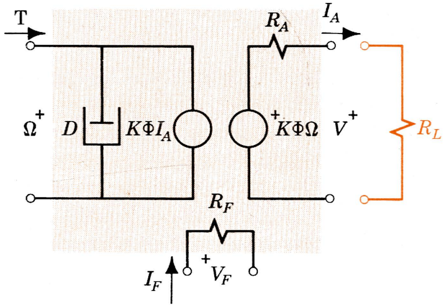


Figure 24.2 Circuit model for a generator.

In the linear model, the mechanical "conductance" term D represents all rotational losses ($P = \Omega^2 D$). It can be evaluated by measuring the no-load power input when the machine is operated at rated speed Ω and rated voltage E (and, therefore, rated air gap flux Φ). Included in the rotational losses are: bearing and brush friction, air friction or *windage,* and core losses due to hysteresis and eddy-currents in regions of changing magnetic flux. (Where is there changing flux in a dc machine?)

The electrical resistance R_A is an effective resistance that takes into account the dc resistance of the armature winding, the effect of nonuniform current distribution in the armature conductors, the brush-contact loss, and the resistance of any auxiliary windings in the armature circuit. It also could include *stray-load* loss that arises from the distortion of flux and current distribution under heavy load. Here we assume that R_A is a fixed, measurable value.

In the model of Fig. 24.2 we assume that D, R_A, R_F, and K are constants; therefore, this is a linear model. To minimize errors, we should determine the constants under rated conditions; for small variations from a chosen operating point, linearity can be assumed with little error. The *external characteristic* is defined by

$$V = E - I_A R_A = K\Phi\Omega - I_A R_A \tag{24-3}$$

Under certain conditions we may choose to assume that the magnetic circuit also is linear. Then the magnetic flux is directly proportional to field current and

$$V = K' I_F \Omega - I_A R_A \tag{24-4}$$

Example 1

A dc generator is rated at 10 kW, 200 V, and 50 A at 1000 rpm. $R_A = 0.4\ \Omega$ and $R_F = 80\ \Omega$. Predict the no-load voltage at 1000 rpm and the full-load voltage at 800 rpm if the field current is kept constant.

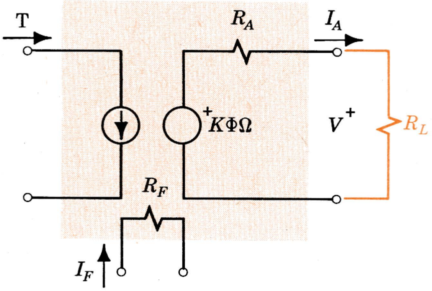


Figure 24.3 Generator model.

For constant-speed operation (resulting from a governor on the driving engine), the mechanical portion of the circuit model is not significant in the calculation of electrical quantities. From the model of Fig. 24.3 we see that the terminal voltage is

$$V = K\Phi\Omega - I_A R_A = E - I_A R_A$$

At no load, $I_A = 0$ and $V = V_{nl} = E$. Therefore

$$V_{nl} = E = V_{fl} + I_A R_A = 200 + 50 \times 0.4 = 220\ \text{V}$$

If the field current is constant, the generated emf E is directly proportional to speed; therefore, at 800 rpm

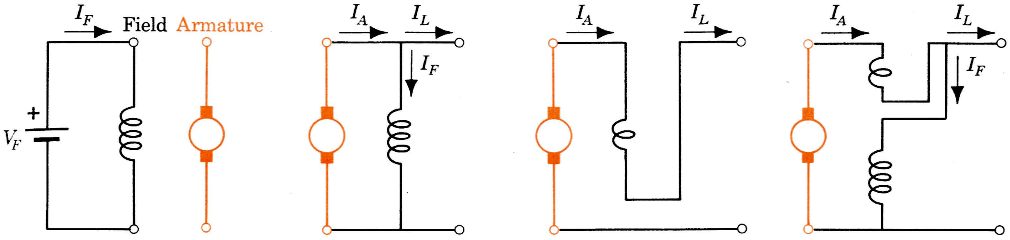
$$E' = E \times \frac{800}{1000} = 220 \times \frac{800}{1000} = 176\ \text{V}$$

The terminal voltage at 800 rpm under full load (50 A) is

$$V'_{fl} = E' - I_A R_A = 176 - 50 \times 0.4 = 156\ \text{V}$$

Field Excitation

One basis for classifying dc machines is according to the arrangement for field excitation. In general, motors and generators may be *separately excited* or *self-excited.* The conventional symbols for field and armature are shown in Fig. 24.4a; here the field is separately excited by connecting it to an independent source. A field winding for *shunt* connection directly across the armature usually consists of many turns of fine wire, and I_F is a few percent of I_A. A field winding for *series* connection consists of a few turns of heavy wire since it carries the entire armature current. Series generators are unsatisfactory for most applications, but series motors are widely used.



(a) Separately excited (b) Shunt connected (c) Series connected (d) Compound connected

Figure 24.4 Wiring diagrams showing possible excitation connections for a generator.

In *compound*-connected machines all field poles have both shunt and series turns. The total mmf is either the sum or difference of $N_{sh}I_F$ and $N_{ser}I_A$ as desired. The relative influence of the shunt windings can be adjusted by means of a variable resistor or *control rheostat* R_C in the shunt field circuit. The series field mmf is a function of load current, but its relative effect can be reduced by connecting a low-resistance *diverter* R_D across the series field. This low-resistance path carries part of the armature current around the series winding.

The relation between generated emf and field current is defined by a *magnetization curve* (Fig. 24.5). Since E is proportional to Φ and I_F is proportional to $\mathfrak{F}$, this is similar to a Φ–$\mathfrak{F}$ curve for a magnetic circuit with an air gap. It is easily obtained experimentally by measuring no-load voltage as a function of field current at a constant speed. Curves for any other speed are obtained by direct proportion. Because of *residual magnetism,* a small emf is generated even with zero field current. Over a wide range the curve is nearly linear. At high values of I_F corresponding to high values of H, *saturation* of the iron core has an appreciable effect.

The Self-Excited Shunt Generator

The magnetization curve is independent of the source of I_F. In a separately excited generator, I_F is controlled by the operator and, within limits, the no-load voltage can be set to any value. In a self-excited generator, the no-load voltage is determined by the magnetization curve and the total resistance of the field

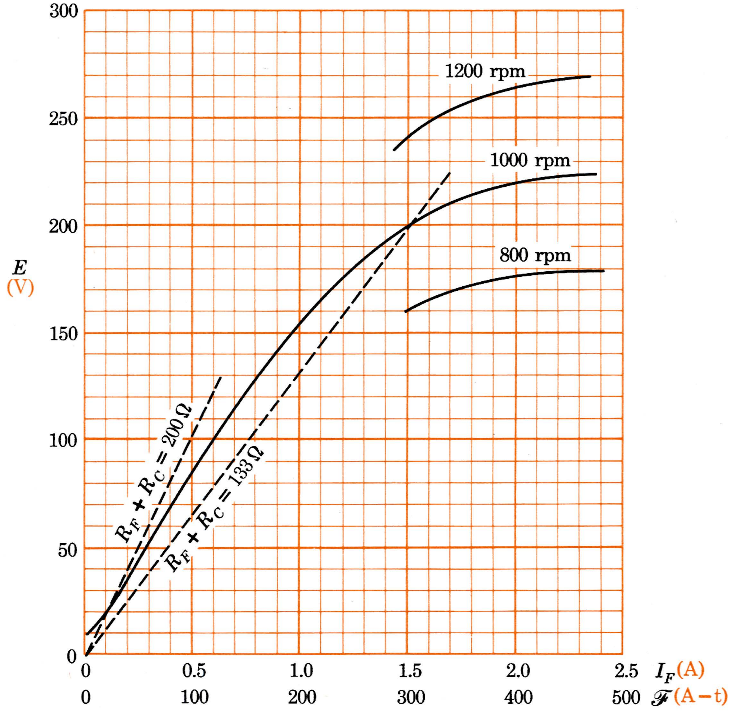
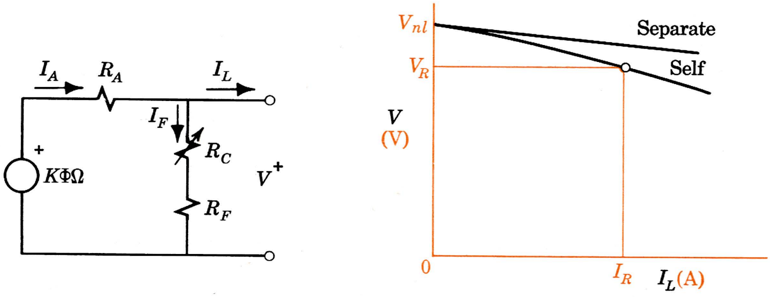


Figure 24.5 Typical magnetization curve.

circuit. (In Fig. 24.6b, V_R and I_R are the *rated* values of terminal voltage and current.) If the very small I_AR_A drop is neglected, the generated emf $E = K\Phi\Omega$ appears across R_F plus R_C, the resistance of the control rheostat, and

$$E = (R_F + R_C)I_F \tag{24-5}$$


I_A
R_A
I_L
I_F
R_C
+
$K\Phi\Omega$
V^+
R_F
V_{nl}
V_R
Separate
Self
V
(V)
0
I_R
I_L(A)

(a) Circuit model

(b) External characteristics

Figure 24.6 A self-excited generator.

This represents a straight line with a slope $R_F + R_C$. But E is also a function of I_F as defined by the magnetization curve. The intersection of these curves is the simultaneous solution of the two relations.

Example 2

The dc generator of Example 1 is rated at 10 kW, 200 V, and 50 A at 1000 rpm. $R_A = 0.4\ \Omega$ and $R_F = 80\ \Omega$. The magnetization curve is shown in Fig. 24.5. Predict the values of R_C to yield a no-load voltage of 200 V and a full-load voltage of 200 V.

For a speed of 1000 rpm, $E = V_{nl} = 200$ V when $I_F = 1.5$ A; therefore, solving Eq. 24-5 yields

$$R_C = \frac{E}{I_F} - R_F = \frac{V_{nl}}{I_F} - R_F = \frac{200}{1.5} - 80 = 53\ \Omega$$

For a full-load voltage of 200 V, the generated emf (see Example 1) must be 220 V. Figure 24.5 indicates a field current of 2 A; therefore,

$$R_C = \frac{V_{fl}}{I_F} - R_F = \frac{200}{2} - 80 = 20\ \Omega$$

The process of voltage buildup in a self-excited generator can be explained in terms of Fig. 24.5. With the generator rotating, a small voltage (about 10 V here) exists due to residual magnetism. When the field circuit is closed, this voltage causes a small field current that increases the flux that increases the voltage, etc. This cumulative process continues until the stable operating point is reached; at this point the voltage generated is just that required to produce the current to establish the necessary flux. No further increase in voltage is possible, except by lowering the resistance of the field circuit. Note that if the field circuit resistance is increased slightly, to 200 Ω in Example 2, the no-load terminal voltage is drastically reduced. (What would happen if the polarity of the residual voltage was opposite to that assumed here?)

The Compound Generator

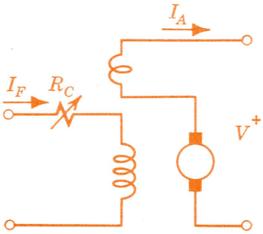
The variation in terminal voltage with a change in load current is usually undesirable. As indicated in Fig. 24.6b, the drop in voltage in a separately excited generator is primarily due to $I_A R_A$. In a self-excited generator the drop in terminal voltage decreases the field current and reduces the generated voltage still further. In contrast, the excitation provided by a series winding increases with armature current. A compound generator with the shunt and series windings properly proportioned produces a full-load voltage just equal to the no-load value; such a machine is said to be *flat-compounded.* (See Example 3 on the following page.)

Under certain circumstances the "drooping" characteristic of the differentially compounded generator may be desirable (Fig. 24.7b). The degree of compounding can be reduced by use of a diverter. Under other circumstances a rise in voltage with load may be desirable and the machine is *over-compounded.*

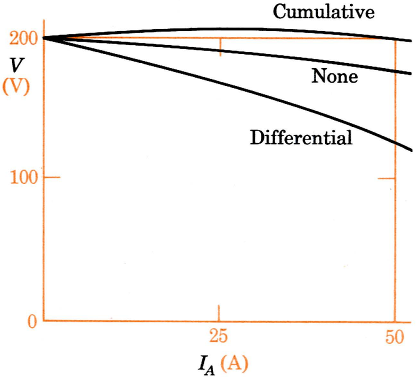
Example 3

The generator of Examples 1 and 2 is to be operated with the shunt field (200 turns per pole) separately excited.

Design a series winding (Fig. 24.7a) to provide flat compounding, and predict the full-load terminal voltage if the series winding is incorrectly connected.



(a) Wiring diagram



(b) External characteristics

Figure 24.7 A compound generator.

We assume that the resistance of the series winding is negligible. From Example 2 (or reading from Fig. 24.5), the required mmf is:

For 200 V at no-load,

$$NI_F = 200 \times 1.5 = 300 \text{ A-t/pole}$$

For 200 V at full-load,

$$NI_F = 200 \times 2 = 400 \text{ A-t/pole}$$

The difference of 400 − 300 = 100 A-t is to be supplied by an armature current of 50 A; therefore, the series winding should provide

$$\frac{100 \text{ A-t/pole}}{50 \text{ A}} = 2 \text{ turns/pole}$$

Knowing the number of turns per pole on the shunt winding used in determining the magnetization curve, we can convert the horizontal scale of Fig. 24.5 to mmf. In *cumulative* compounding the mmfs of the shunt and series windings add. In *differential* compounding the series winding is connected so that the series mmf is in opposition to that produced by the shunt field.

For differential compounding with $I_F = 1.5$ A and $I_A = 50$ A, the net mmf per pole is

$$200 \times 1.5 - 2 \times 50 = 300 - 100 = 200 \text{ A-t}$$

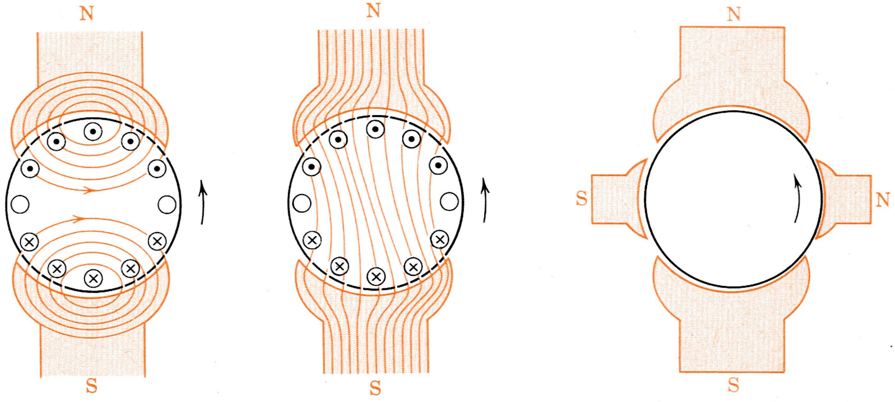
This corresponds to $I_F = 1$ A (in the 200-turn shunt winding) and the magnetization curve indicates $E \cong 156$ V. Then

$$V = E - I_A R_A = 156 - 50 \times 0.4 = 136 \text{ V}$$

The external characteristics of the generator are as shown.

Armature Reaction

In deriving the basic relations for a conductor moving in a magnetic field, we assumed that the field due to current in the conductor itself was negligibly small. In a loaded machine, this is no longer true. As shown in Fig. 24.8a, the armature current establishes an appreciable magnetic field directed at right angles or *in quadrature* with respect to the main field. The result of this *armature reaction* is a distortion of the flux distribution (Fig. 24.8b) with two practical results.



(a) Quadrature field (b) Combined field (c) Commutating poles

Figure 24.8 Flux distribution showing armature reaction in a generator.

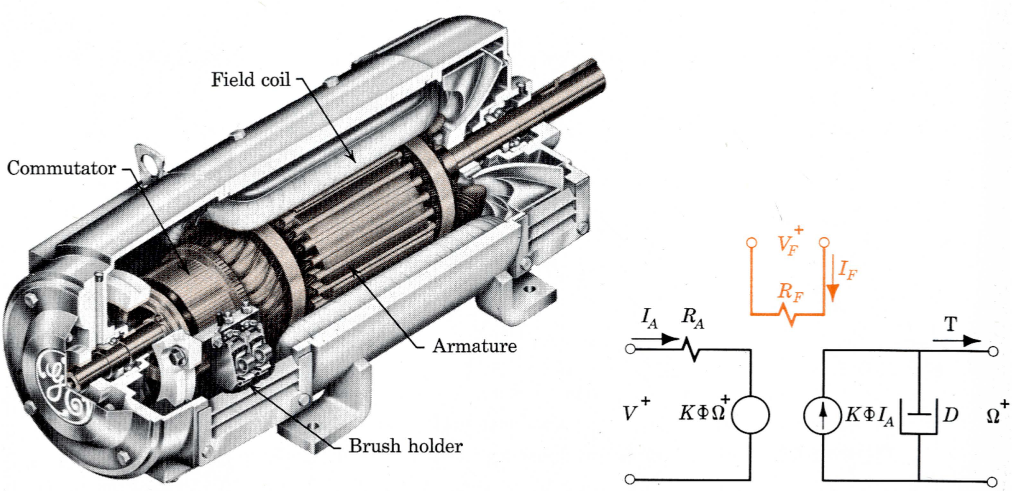
In the first place, commutation is affected adversely because the point of zero generated emf is shifted (in the direction of rotation in a generator). One practical arrangement is to place conductors carrying armature current in the faces of the main poles. Such *compensating windings* oppose the armature mmf and reduce the distortion of the flux distribution. Most machines also have small auxiliary field windings carrying the load current and placed midway between the main poles. Such *commutating poles* aid in reversing the current in the conductors as they undergo commutation by providing the necessary voltage impulse ($v \, \Delta t = L \, \Delta i$).

A second effect of armature reaction is a reduction in generated emf due to a reduction in the average flux density. Because the magnetic circuit is nonlinear, the increase in flux density in regions of high mmf is less than the decrease in regions of low mmf. The total flux per pole is reduced and therefore the emf. This effect contributes to the curvature in the external characteristics in Fig. 24.7b.

Armature reaction is of great importance in the design and performance of large machines, and certain devices such as the automobile generator voltage regulator depend on armature reaction for their operation. In most of our problems, however, the additional complexity of considering armature reaction is not justified and we shall neglect its effect.

Direct-Current Motors

If electrical power is supplied to a dc machine, it can operate as a motor. The essential difference between a motor and a generator is in the direction of armature-current flow. The same circuit model is applicable to a motor, or it can be turned around as in Fig. 24.9. DC motors are classified as shunt, series, or compound, according to the method of field connection. Since torque depends



(a) Cutaway view (b) Circuit model

Figure 24.9 A dc motor. (Courtesy General Electric Company.)

on field flux and armature current, the operating characteristics of the various types differ widely.

The Shunt Motor

The basic relations (Eqs. 24-1) apply to the shunt motor, but it is usually more convenient to work with speed in rpm. Letting $\Omega = n \times 2\pi/60$, the basic equations are:

$$E = K\Phi\frac{2\pi n}{60} = kn\Phi \qquad \text{and} \qquad \mathrm{T}_d = K\Phi I_A \tag{24-6}$$

where n is the speed in rpm and k is a new constant.

On the basis of Fig. 24.9b, the voltage equation becomes

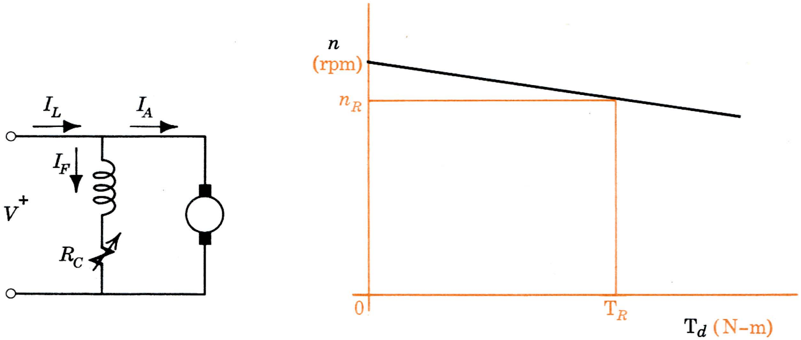
$$V = E + I_A R_A = kn\Phi + I_A R_A \tag{24-7}$$

The speed characteristic is revealed by solving Eq. 24-7 so that

$$n = \frac{V - I_A R_A}{k\Phi} \tag{24-8}$$

With Constant Voltage. In Fig. 24.10, the field circuit is connected directly across the supply line; the armature voltage is constant and the field current is constant for any setting of the rheostat R_C. Except for the effect of armature reaction, Φ remains constant. Since even at full load $I_A R_A$ is a small percentage of V, the shunt motor is essentially a constant-speed machine. At any given load, however, the speed is readily adjusted, within limits, by changing the field current with a control rheostat.

A specific expression for speed as a function of developed torque is obtained



(a) Wiring diagram (b) Speed–torque characteristic

Figure 24.10 A shunt motor with constant applied voltage.

by solving the torque equation for I_A and substituting in Eq. 24-8 to yield

$$n = \frac{V}{k\Phi} - \frac{R_A}{kK\Phi^2}T_d = n_{nl} - m\,T_d \tag{24-9}$$

Equation 24-9 indicates a straight-line relation between speed and developed torque (Fig. 24.10b).

Example 4

The dc machine of the previous Examples is to be used as a shunt motor (Fig. 24.11). No-load tests indicate that the rotational losses at rated speed are 1100 W. For a terminal voltage of 220 V, specify the field current for a full-load ($I_A = 50$ A) speed of 1000 rpm, and estimate the no-load speed.

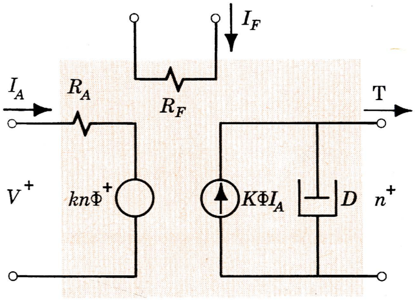


Figure 24.11 Shunt motor calculations.

Solving Eq. 24-7, for $I_A = 50$ A and $R_A = 0.4\ \Omega$,

$$E = V - I_A R_A = 220 - 50 \times 0.4 = 200 \text{ V}$$

The magnetization curve at a full-load speed of 1000 rpm (Fig. 24.5) indicates $I_F = 1.5$ A for $E = 200$ V.

To supply the rotational losses, the armature input must be 1100 W (plus negligibly small $I_A{}^2R_A$ loss). Therefore, at no load,

$$I_A = \frac{P_{nl}}{V} = \frac{1100}{220} = 5 \text{ A}$$

The no-load speed could be determined by solving Eq. 24-8, using the value of $k\Phi$ determined from full-load data. A simpler approach is to use the proportionality between E and n which leads to

$$n_{nl} = \frac{E_{nl}}{E_{fl}} \times n_{fl} = \frac{(V - I_A R_A)_{nl}}{(V - I_A R_A)_{fl}} \times n_{fl}$$

$$= \frac{218}{200} \times 1000 = 1090 \text{ rpm}$$

With Adjustable Voltage. Greater flexibility is possible if the armature and field voltages are obtained from separate power supplies. As indicated by Eq. 24-8, speed can be controlled by armature voltage V or field voltage V_F (actually Φ). By holding Φ constant, the torque capability can be maintained (Eq. 24-6) while the speed can be effectively controlled over a wide range by adjusting the armature voltage. This arrangement is commonly used in industrial applications of large motors.

Just how does a machine adjust to an increased load? As torque is applied to the shaft the motor slows down and the emf decreases. A small decrease in emf results in a large increase in I_A (Eq. 24-7) and a large increase in developed torque. The motor slows down just enough to develop the required torque. If the shaft load is reduced, the motor speeds up slightly and the armature current (representing power input) is reduced.

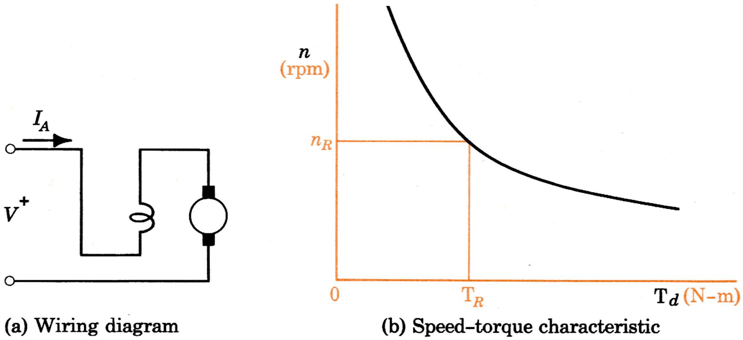
As indicated by Example 4, the variation in speed of a constant-voltage shunt motor under load is quite small. The *speed regulation,* analogous to the voltage regulation of a transformer or generator, is defined as

$$\mathrm{SR} = \frac{n_{nl} - n_{fl}}{n_{fl}} \tag{24-10}$$

For the shunt motor of Example 4, the speed regulation is 10%. Typical values for shunt motors are in the range from 5 to 12% and justify classifying the shunt motor as a constant-speed machine.

The Series Motor

If a field winding consisting of a few turns of heavy wire is connected in series with a dc armature (Fig. 24.12), the resulting *series motor* has characteristics quite



(a) Wiring diagram (b) Speed–torque characteristic

Figure 24.12 A series motor.

different from those of a shunt motor. The basic equations (Eq. 24-6) still apply, but now magnetic flux is dependent on load current.

To simplify the analysis, we shall assume that the series motor is operating on the linear portion of the magnetization curve and $\Phi = k_1 I_A$. Then the emf and torque equations become

$$E = kn\Phi = knk_1 I_A = k_E n I_A \tag{24-11a}$$

and

$$\mathrm{T}_d = K\Phi I_A = Kk_1 I_A{}^2 = k_\mathrm{T} I_A{}^2 \tag{24-11b}$$

Equation 24-11b indicates that the torque developed by a series motor is proportional to the square of the armature current. (How does this compare with the performance of a shunt motor?) To determine the speed-torque characteristic, we note that

$$V = E + I_A(R_A + R_{\mathrm{ser}}) = (k_E n + R_T)I_A \tag{24-12}$$

where $R_T = R_A + R_{\mathrm{ser}}$, the total resistance of the armature circuit. Solving Eq. 24-12 for I_A and substituting in Eq. 24-11b,

$$\mathrm{T}_d = k_\mathrm{T} \frac{V^2}{(k_E n + R_T)^2} \tag{24-13}$$

For values of n such that $k_E n \gg R_T$, Eq. 24-13 indicates that the developed torque is inversely proportional to the square of the speed. This relation is displayed in Fig. 24.12b. The series motor is *not* a constant-speed machine; this is not necessarily a disadvantage, but it does limit the application of series motors. In Example 5, if the torque drops to approximately 30% of the full-load value, the machine speed doubles; to prevent excessive speeds, series motors are mechanically coupled to their loads at all times. (See Example 5 on the next page.)

Starting

At starting the speed is zero, the emf of a dc machine is zero, and the applied voltage appears across the armature resistance (see Eq. 24-7 or Fig. 24.9). To prevent excessive armature current, the starting voltage must be reduced. If the armature voltage is adjustable, the motor is smoothly accelerated by increasing the voltage from zero. Automatic control to limit the starting current to 1.5 or 2 times rated is often used. If power is supplied at constant voltage, a starting resistance is connected in series with the armature; the resistance is shorted out in one or several steps as the motor comes up to speed. For high starting torque in either case, full voltage should be applied to the field of a shunt motor.

In a series motor, the field current is equal to the armature current. Since

Example 5

(a) The dc machine of the previous Examples is to be operated as a series motor. The resistance of the series winding is estimated to be 0.2 Ω. Design the series winding.

(b) Predict the torque developed at rated speed and at twice rated speed.

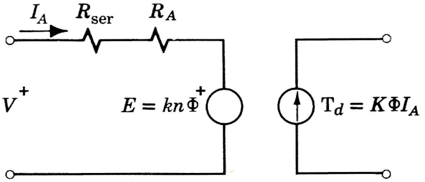


Figure 24.13 Series motor performance calculations using circuit model.

Assuming that the machine is to be rated at 220 V and 50 A at 1000 rpm, at full load the emf must be

$$E = V - I_A(R_A + R_{ser}) = 220 - 50(0.4 + 0.2) = 190 \text{ V}$$

The magnetization curve (Fig. 24.5) indicates a required mmf of $NI = 200 \times 1.4 = 280$ A-t. Then the series winding must provide

$$N_{ser} = \frac{NI}{I_A} = \frac{280 \text{ A-t}}{50 \text{ A}} \cong 5\tfrac{1}{2} \text{ turns per pole}$$

If this is not convenient, 6 turns can be wound per pole and a diverter used to ensure 280 A-t.

To determine torque we must evaluate two machine constants. We assume operation on the linear portion of the magnetization curve and use the linear model of Fig. 24.13. (Is this a reasonable assumption for currents lower than rated? See Fig. 24.5.) From Eq. 24-6, using rated values,

$$K\Phi = \frac{60E}{2\pi n} = \frac{60 \times 190}{2\pi \times 1000} = \frac{5.7}{\pi}$$

and the rated value of developed torque is

$$T_R = K\Phi I_A = \frac{5.7}{\pi} \times 50 \cong 91 \text{ N-m}$$

From Eq. 24-11a,

$$k_E = \frac{E}{nI_A} = \frac{190}{1000 \times 50} = 0.0038$$

Equation 24-13 indicates that at twice rated speed,

$$T_2 = T_R \frac{(k_E n + R_T)_R{}^2}{(k_E n + R_T)_2{}^2} = 91\left(\frac{3.8 + 0.6}{7.6 + 0.6}\right)^2 \cong 26 \text{ N-m}$$

the torque varies as the square of the current (neglecting saturation), for a starting current higher than rated value the starting torque of a series motor is greater than that of a comparable shunt motor.

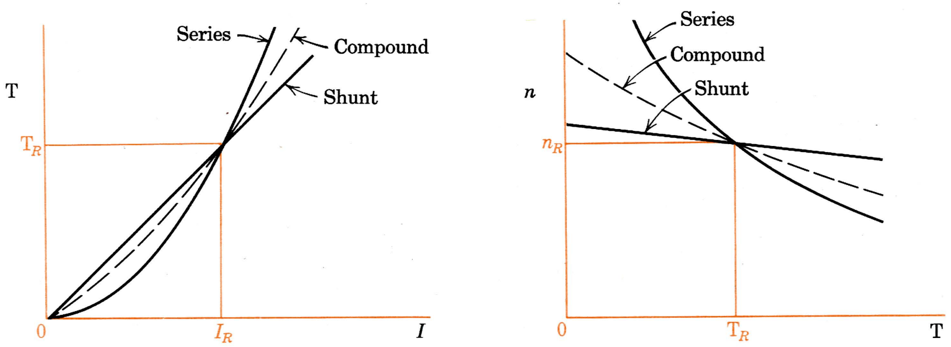
Special *starting boxes* for dc motors provide a variable starting resistance and protection against loss of voltage and loss of field current. (See Exercise 16.) Reversing either the armature or the field connection, but not both, changes the direction of rotation.

Speed Control

The shunt motor offers two important virtues: ease in adjusting speed and flexibility in matching the speed-torque characteristics of the load. To match the constant-torque characteristic of many mechanical loads, a motor can be operated with rated field and armature currents and adjustable armature voltage; developed torque is then independent of speed. Other loads require high torque at low speed and low torque at high speed; such "constant-horsepower" characteristics can be matched by using constant armature voltage and varying the field current. Because of magnetic circuit saturation in a practical machine, the range of speed control by field current is limited. The lowest speed corresponds to the maximum field current when $R_C = 0$. (What would happen if the field circuit were opened accidentally?) Large machines may employ both methods of speed control by using sophisticated regulators on both armature and field circuits.

The speed of a series motor is dependent on the load and is not so easily controlled. Under heavy torque loads, the series motor slows down, the emf decreases, and the armature current increases until sufficient torque is developed. An adjustable voltage supply or a series resistor can be used to vary the speed. (What would happen if the mechanical load were disconnected accidentally?) The high starting torque of series motors makes them useful in locomotives, trolley buses, cranes, and hoists.

The *compound motor* has a shunt and a series field connected so that their mmfs add. The operating characteristics are intermediate between those of shunt and series motors. (See Fig. 24.14.) By adjusting the relative strength of the shunt and series fields, a motor can be designed to provide a desired compromise.



(a) Torque–current characteristics (b) Speed-torque characteristics

Figure 24.14 Operating characteristics of dc motors.

Adjustable Voltage Supply

To realize all the virtues of a dc motor, an adjustable voltage power supply is required.

Ward-Leonard System. One possibility is to use an ac motor driving a dc generator and an *exciter* (to supply field excitation current). In Fig. 24.15, the motor-generator set converts ac power to adjustable voltage dc power. With control rheostat R_{CM} set for rated field current, dc motor voltage V is controlled by the generator field current through R_{CG}. Motor speed can be controlled smoothly from standstill to rated speed in either direction; reversing the direction of generator field current reverses the direction of motor rotation. Speeds above rated can be obtained by reducing the motor field current. *Regenerative braking* of the load is possible by reducing the generator field current thereby reducing V below the motor emf $E = kn\Phi$ and reversing the power flow; the generator acts as a motor driving the ac machine as a generator feeding the rotational energy of the load back into the ac line. Regenerative braking saves energy while providing smooth control.

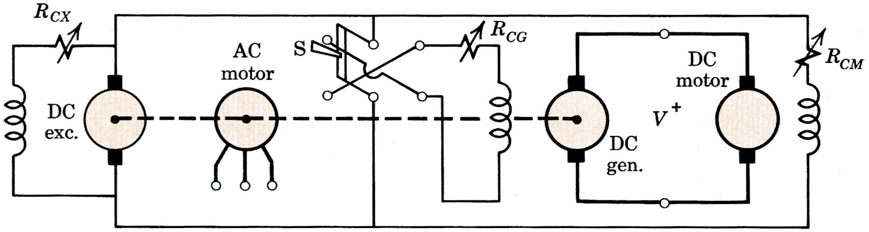


Figure 24.15 Ward-Leonard system adjustable voltage supply.

Thyristor Rectifier System. Electronic rectifiers using SCRs offer several significant advantages over motor-generator sets. SCRs are more efficient, longer lived, lighter in weight, and more easily maintained; they do not, however, offer the inherent regenerative braking capability of an m-g set.

The half-wave SCR supply of Fig. 24.16 is suitable for driving small dc motors. The dc field current is provided through diode D_1 from the ac supply; when the ac polarity reverses cutting off D_1, field current continues to flow through *free-wheeling diode* D_2. (The inductance of the field itself provides some filtering action.) Capacitor C is charged through D_3 and an adjustable reference voltage V_{ref} is obtained at the speed-controlling potentiometer P. Whenever the tachometer voltage differs from the reference voltage, i.e., whenever the motor speed

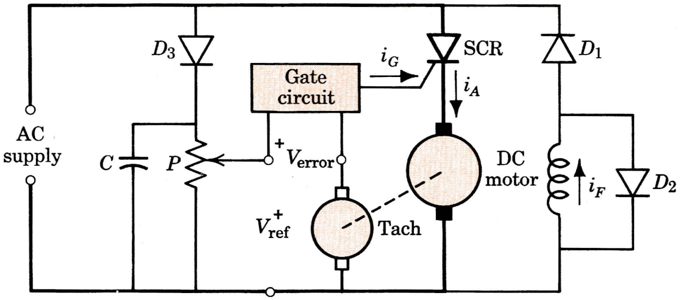

D3
i_G
SCR
D1
Gate circuit
i_A
AC supply
C
P
+
V_{error}
DC motor
i_F
D2
+
V_{ref}
Tach

Figure 24.16 Half-wave SCR adjustable voltage supply.

differs from the reference speed, the resulting error voltage shifts the firing angle of the gate circuit and thereby adjusts the conduction angle of the SCR that supplies armature current. (See also Fig. 13.29.)

Performance of DC Motors

Previous examples have illustrated the approach to calculating speed and torque. To compute output and efficiency, the mechanical and electrical losses must be taken into account. The parameters in the circuit model may be computed by using rather elaborate formulas, or they may be measured in the laboratory. A no-load test of a rotating machine is similar to the open-circuit test of a transformer and yields analogous data. Full-load tests may be performed, using a *dynamometer* to measure the shaft torque.

The *power-flow diagram* in Fig. 24.17 is an aid in visualizing the energy conversion processes in a machine. The power-flow diagram, the energy balance, and the circuit model are closely related. Under dynamic conditions additional terms must be introduced to account for changes in magnetic and mechanical energy storage. The midpoint of the diagram corresponds to the air gap with iron losses entered on the mechanical side.

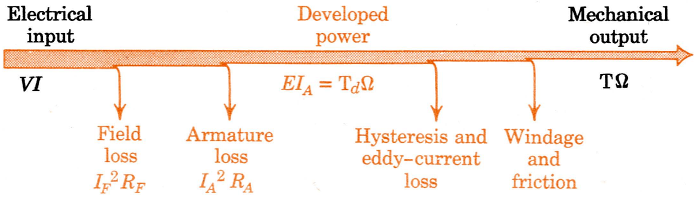

Electrical input
Developed power
Mechanical output
VI
$EI_A = T_d\Omega$
$T\Omega$
Field loss
$I_F^2 R_F$
Armature loss
$I_A^2 R_A$
Hysteresis and eddy-current loss
Windage and friction

Figure 24.17 Power-flow diagram for a dc motor.

Example 6

The dc machine of Example 4 is rated as a shunt motor at 220 V, 50 A, at 1000 rpm. $R_F = 80\ \Omega$ and $R_A = 0.4\ \Omega$. At rated speed, the field current is 1.5 A and the rotational losses are 1100 W. Determine the output under rated conditions and predict the full-load efficiency. (Do not charge $I_F{}^2R_C$ losses to the motor.)

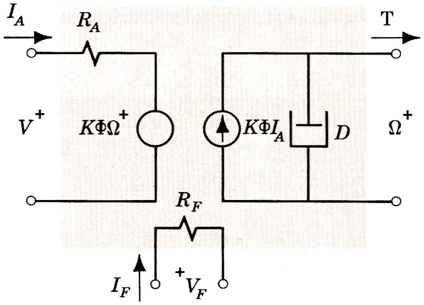


Figure 24.18 Shunt motor model.

The circuit model of Fig. 24.18 is applicable. At rated current and voltage,

$$E = V - I_AR_A = 220 - 50 \times 0.4 = 200\text{ V}$$

By Eq. 24-6,

$$K\Phi = \frac{60E}{2\pi n} = \frac{60 \times 200}{2\pi \times 1000} = \frac{6}{\pi}$$

and the rated value of developed torque is

$$\mathrm{T}_d = K\Phi I_A = \frac{6}{\pi} \times 50 = \frac{300}{\pi} = 95.5\text{ N-m}$$

The developed mechanical power at the air gap is

$$P_M = \mathrm{T}_d\Omega = K\Phi I_A\Omega = \frac{300}{\pi}\frac{2000\pi}{60} = 10{,}000\text{ N-m/s}$$

Alternatively, the developed mechanical power is equal to the electrical power at the air gap or

$$P_M = P_E = EI_A = 200 \times 50 = 10{,}000\text{ W or N-m/s}$$

Subtracting the rotational losses, the power available at the shaft is

$$P = 10{,}000 - 1100 = 8900\text{ W or } \frac{8900}{746} = 11.9\text{ hp}$$

The overall efficiency is

$$\frac{\text{Output}}{\text{Input}} = \frac{\text{shaft output}}{VI_A + I_F{}^2R_F} = \frac{8900}{220 \times 50 + 1.5^2 \times 80}$$

$$= \frac{8900}{11{,}180} \cong 0.8 \text{ or } 80\%$$

Although dc motors are less common than ac motors, they are ideally suited to certain applications. The commutator makes the dc machine relatively high in first cost and in maintenance, and where there is moisture, dust, or explosive vapor, the enclosure must be selected with care. However, the dc motor provides easily adjusted speed, high efficiency, and great flexibility. DC motors range in size from tiny units turning fans in electronic apparatus to 10,000-hp motors driving the rolls in steel mills. Other applications include ship propulsion, fork lifts, punch presses, and battery-powered vehicles.

SUMMARY

- In a direct-current machine, the basic steady-state relations are:

$$E = K\Phi\Omega \qquad \text{and} \qquad \mathrm{T}_d = K\Phi I$$

- In a dc generator, the external characteristics are defined by:

$$V = E - I_A R_A \qquad \text{and} \qquad \mathrm{T}_d = \mathrm{T}_a - \Omega D$$

  Generators may be separately excited or self-excited.
  The magnetization curve defines E as a function of I_F.

- Self-excited voltage buildup stabilizes where the field-resistance line intersects the magnetization curve.
  In a flat-compounded generator the increase in series field mmf $N_{\text{ser}} I_A$ offsets the drop in armature voltage $R_A I_A$.
  Armature reaction results in a distortion of flux distribution.

- A dc motor differs from a generator only in the direction of I_A;

$$E = kn\Phi = V - I_A R_A \qquad \mathrm{T}_d = K\Phi I_A = \mathrm{T}_a + \Omega D$$

	SHUNT MOTOR	SERIES MOTOR
Flux:	Constant but adjustable	Varies with I_A
Emf:	$E = kn\Phi$	$E = k_E n I_A$
Torque:	$\mathrm{T}_d = K\Phi I_A$	$\mathrm{T}_d = k_\mathrm{T} I_A^2$
Speed:	Nearly constant $n = \dfrac{V - I_A R_A}{k\Phi}$ $n = n_{nl} - k'\mathrm{T}_d$	Widely varying $\mathrm{T}_d = k_\mathrm{T} \dfrac{V^2}{(k_E n + R_T)^2}$ $\dfrac{n}{n_R} \cong \left(\dfrac{\mathrm{T}_{dR}}{\mathrm{T}_d}\right)^{1/2}$
Speed Control:	Efficient over wide range, by varying V_A and I_F	Some control by varying V_A or inserting R_{ser}.
Starting Torque:	Fair; $\dfrac{\mathrm{T}}{\mathrm{T}_R} = \dfrac{I_A}{I_{AR}}$	Good; $\dfrac{\mathrm{T}}{\mathrm{T}_R} = \left(\dfrac{I_A}{I_{AR}}\right)^2$

In starting a dc motor, $E = 0$ at $n = 0$; therefore, must start with low V or insert series R to limit I_A and $I_A^2 R_A$ heating.

Speed regulation of a motor is defined as $\mathrm{SR} = \dfrac{n_{nl} - n_{fl}}{n_{fl}}$.

Characteristics of compound motors are intermediate between those of shunt and series motors.

REVIEW QUESTIONS

1. Explain the statement: "The air-gap power represents the reversible portion of electromechanical energy conversion."
2. Outline a laboratory procedure for determining D in Fig. 24.1.
3. Why and where are there eddy-current and hysteresis losses in a dc machine?
4. Given the rated voltage of a generator, estimate the no-load voltage.
5. On a compound generator, how could you distinguish between the series and shunt field windings by inspection?
6. Draw a wiring diagram of a compound generator showing control rheostat and diverter; what is the function of each?
7. What would happen if the polarity of the residual magnetism in a self-excited generator were reversed?
8. Explain the operation of a flat-compounded generator.
9. Why is an undercompounded generator more stable than an overcompounded machine? Sketch their V-I characteristics.
10. What is armature reaction and why is it of concern to designers?
11. Sketch speed-torque curves for series, shunt, and compound motors.
12. What happens if the field connection of a fully loaded shunt motor is accidentally broken? Repeat for a lightly loaded motor.
13. How does a shunt motor react (electrically) to increased load?
14. How does a series motor react (electrically) to increased load?
15. In starting a shunt motor, should V_F be maximum or minimum? Why?
16. In comparison to I_F control of speed, why is V_A control more expensive and R_A control less efficient?
17. What are the advantages of an SCR voltage supply over an m-g set?
18. How is the output voltage of an SCR supply adjusted?

EXERCISES

1. A separately excited dc generator is rated at 250 V, 100 A, 2000 rpm; $R_A = 0.2\,\Omega$, $R_F = 125\,\Omega$.
   (a) Draw an appropriate linear model for steady-state, constant-speed operation.
   (b) Predict the no-load terminal voltage ($I_A = 0$).
   (c) Predict the full-load terminal voltage ($I_A = 100$ A) if the field current is decreased 10%.
2. A 20-kW, 250-V, 1200-rpm generator is driven at a constant speed of 1200 rpm and the field current is adjusted to provide an output voltage of 250 V at a load current of 80 A. When the load is disconnected, the output voltage rises to 262 V. Evaluate the effective armature resistance.
3. When the load on the generator of Exercise 2 is disconnected, the no-load driving power is 800 W.
   (a) Evaluate the D factor in Fig. 24.2 for this generator.
   (b) Predict the driving torque required under full load.
4. The generator of Exercises 2 and 3 is running without load ($I_A = 0$). If the speed is increased to 1500 rpm and the field current reduced by 10%, predict the new terminal voltage and the new driving power.
5. The magnetization curve of Fig. 24.19 describes a separately excited dc generator. Predict the no-load voltage at:
   (a) 1200 rpm with $I_F = 2$ A;
   (b) 2100 rpm with $I_F = 1$ A.

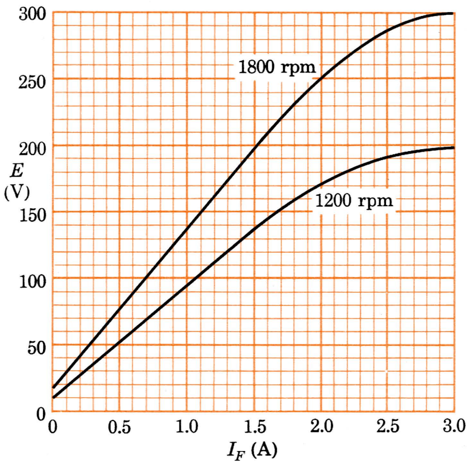


Figure 24.19 Magnetization curve.

6. A separately excited dc generator follows the curve of Fig. 24.19. Specify the field current for a no-load voltage of 150 V at:
   (a) 1800 rpm.
   (b) 1500 rpm.
7. The magnetization curve for a dc generator is shown in Fig. 24.19. The generator runs at 1800 rpm and is rated at 250 V, 40 A. $R_A = 0.25\ \Omega$ and $R_F = 100\ \Omega$.
   (a) Estimate the field current for rated conditions.
   (b) Predict the no-load voltage for separate excitation at 250 V.
8. A self-excited dc generator, known to be in good mechanical condition, fails to generate anywhere near rated voltage when driven by a motor. List four possible causes for this unsatisfactory operation and indicate the steps you would take to eliminate each.
9. A self-excited dc generator operates at 1800 rpm (see Fig. 24.19). Predict the no-load voltage if the total resistance of the field circuit is 120 Ω. Repeat for operation at 1200 rpm.
10. A compound generator with $R_A = 1\ \Omega$ and $R_{\text{ser}} = 0.2\ \Omega$ has 2 turns/pole on the series field and 300 turns/pole on the shunt field. The no-load voltage is 200 V with separate excitation at $I_F = 1$ A. Stating any necessary assumptions:
    (a) Estimate the terminal voltage at $I_A = 30$ A with the same separate excitation.
    (b) Repeat part (a) for compound excitation and describe the degree of compounding.
11. In Fig. 24.20, $V = 200$ V, auxiliary resistor $R = 100\ \Omega$, and $R_F = 100\ \Omega$; R_A is negligibly small.
    (a) Determine the relation between speed and torque for this special machine.
    (b) If the speed is 1200 rpm, predict the developed torque at $I = 1.5$ A.

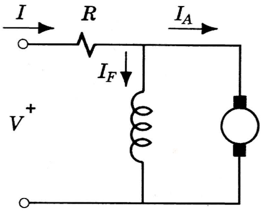


Figure 24.20

12. Predict how I_A and speed n of a typical dc shunt motor would be affected by the following changes in operating conditions:
    (a) Keeping I_F and T constant, and halving V,
    (b) Keeping I_F and P constant, and halving V,
    (c) Keeping V and T constant, and doubling I_F,
    (d) Keeping P constant, and halving I_F and V, where T = load torque, V = terminal voltage, and P = power output.
13. A shunt motor is rated at 10 hp, 230 V, 42 A, and 1500 rpm. $R_F = 115\ \Omega$ and $R_A = 0.25\ \Omega$.
    (a) Estimate the no-load speed.
    (b) Predict the speed when operated at 250 V with line current = 26 A and with I_F adjusted to the rated value of 2 A.

(c) Predict the no-load speed at 230 V with a control rheostat $R_C = 62\ \Omega$ in series with the field.

14. A 100-hp, 700-V series motor is rated at 120 A at 1200 rpm ($R_A = 0.25\ \Omega$, $R_{\text{ser}} = 0.1\ \Omega$). For $I_A = 60$ A, predict the speed and the power output by:
(a) Making simplifying assumptions for an approximate answer.
(b) Using the derived relations (Eqs. 24-12 and 24-13). Draw a conclusion.

15. State how I_A and n (see Exercise 12) of a typical dc series motor would be affected (approximately) by halving the terminal voltage if:
(a) T remains constant.
(b) P remains constant.
(c) T varies as the square of n.

16. Using the starting box of Fig. 24.21, a shunt motor is started by closing switch S and moving contact lever L to the first button on R_1. As the motor speeds up, the lever is advanced; at the last button it is held in position by magnet M against the pull of a spring. Explain the functions of R_1, R_2, and M in terms of desired starting conditions for a dc motor.

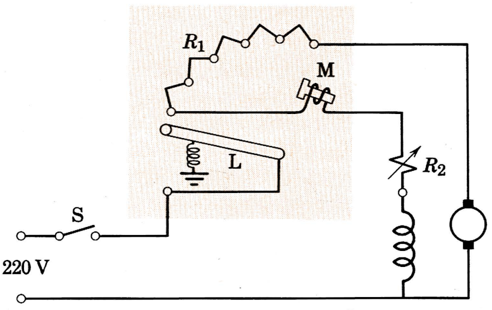


Figure 24.21 Starting box.

17. Design a starting box (i.e., specify the value of R_1 in Fig. 24.21) for the shunt motor of Exercise 13 to limit the starting armature current to twice rated.

18. In starting the motor of Exercise 14, it is desired to limit the starting current to twice-rated current.
(a) Draw a schematic wiring diagram showing a starting resistance and specify the value.
(b) Estimate the starting torque, stating any necessary assumptions.

19. The output of a paper mill is to be wound on a 10-in. diameter core at a constant speed of 115 ft/s with a tension of 120 lb. The finished roll is 30-in. in diameter. Specify the motor to drive the windup roll by hp, type, and method of speed control.

20. A dc shunt motor is rated at 10 hp, 200 V, 1200 rpm. $R_F = 100\ \Omega$ and $R_A = 0.4\ \Omega$. *Stating* any necessary assumptions:
(a) Estimate the no-load speed.
(b) Estimate the no-load speed with a 50-Ω control rheostat in the field circuit.
(c) With the control rheostat of part (b) removed and with a 2-Ω control resistance in series with the armature, estimate the starting current and torque and predict the running speed at $I_A = 40$ A.

21. A series motor ($R_T = 0.4\ \Omega$) draws 50 A at 250 V while hoisting a given weight at 300 ft/min. Assuming linearity, specify the resistance to be placed in series with the armature to slow the hoisting speed to 200 ft/min.

22. The torque required by a mixer is defined by $T = 3 \times 10^{-5} n^2$ newton-meters where n is speed in rpm. Available are a shunt motor ($R_A = 0.4\ \Omega$, $R_F = 240\ \Omega$) and a series motor ($R_T = 0.6\ \Omega$), each rated at 10 hp and 1200 rpm at 240 V.
(a) Sketch the speed-torque curves of the two motors and the mixer on the same graph.
(b) Stating any necessary assumptions, predict the operating speeds of the two motors.

23. A 20-hp, 240-V, 1200-rpm shunt motor with $R_A = 0.2\ \Omega$ follows the magnetization curve of Fig. 24.5.
(a) Estimate the field current for full-load ($I_A = 70$ A).
(b) While running at full load, the field current is quickly increased to 1.7 A. Predict the new I_A and the initial torque available for regenerative braking.

24. A small "personal" electric car for urban use requires full power to accelerate to full speed of 30 km/h in 10 s, but only 30% power to cruise at full speed. A typical cycle consists of acceleration, cruising for 1 minute, and deceleration to a stop. Estimate the percentage increase in operating range possible if regenerative braking at 60% efficiency is used.

25. The ac supply voltage in Fig. 24.16 is $v_s = V_m \sin \omega t$ and the gate circuit fires the SCR at $\omega t = 60°$. On the same time axis:
  (a) Sketch the waveforms of v_s and i_A, taking into account the significant inductance of the armature.
  (b) Show i_G and i_F waveforms, assuming the time constant of the field winding is approximately equal to one cycle of the supply.
  (c) Show the waveform of v_A. (Hint: How is v_A related to e_A during the time the SCR is OFF?)

26. Given a self-excited shunt generator,
  (a) Draw an appropriate steady-state circuit model for this machine.
  (b) Draw a clearly labeled power-flow diagram with input on the left.

27. Calculate the rated efficiency of the series motor of Exercise 14.

28. Calculate the rated efficiency and predict the speed regulation and no-load armature current of the shunt motor of Exercise 13.

29. A 20-hp, 250-V, 1500-rpm shunt motor with $R_A = 0.2\ \Omega$ requires a no-load total input of 8 A.
  (a) Predict the full-load efficiency.
  (b) Predict the speed regulation.

30. A shunt motor with automatic controls is designed for a "base speed" of 1000 rpm at which it delivers 100% torque and 100% horsepower at full armature voltage and full field current. Above or below base speed, control is by adjusting I_F or V_A. Plot torque and horsepower (for rated I_A) vs. speed from 0 to 2000 rpm. Label the regions controlled by I_F and V_A.

PROBLEMS

1. A shunt generator is rated at 100 V at 1 A with $I_F = 0.02$ A; another is rated at 100 V at 25 A with $I_F = 1$ A. Show how these could be connected as a two-stage power amplifier. Estimate the dc power gain possible.

2. The magnetization curve for a 1800-rpm shunt generator is shown in Fig. 24.19. Machine constants are: $R_A = 0.5\ \Omega$, $L_A = 0.5$ H, $R_F = 120\ \Omega$, and $L_F = 30$ H. For self-excited operation:
  (a) Estimate the steady-state no-load terminal voltage.
  (b) Predict the time required for the voltage to build up to 80% of its ultimate no-load value. (Hint: What voltage is available to cause di/dt?)

3. A 100-hp, 700-V shunt motor with $R_A = 0.3\ \Omega$ and $R_F = 250\ \Omega$ operates at 2000 rpm with a control rheostat of 50 Ω in the field circuit. Predict the no-load speed with and without the 50-Ω rheostat.

4. A series motor ($R_T = 0.25\ \Omega$) is rated at 25 hp at 1500 rpm with 85 A at 250 V. Calculate the rated output torque and predict the current, speed, and hp output if the load torque is twice the rated value.

5. Predict the steady-state performance of a machine labeled: DC SHUNT MOTOR, 50 hp, 1500 rpm, 250 V, 180 A. At no load, rated speed requires $I_F = 3$ A and $I_A = 15$ A.
  (a) Draw a clearly labeled schematic wiring diagram, an appropriate circuit model, and a power-flow diagram.
  (b) Mechanical load is applied to the shaft

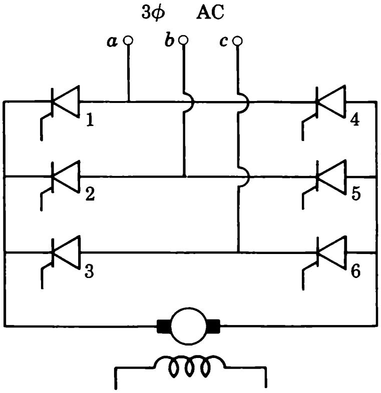


Figure 24.22

until the total line current is 143 A. Predict the speed, developed torque, mechanical power output in hp, and the overall efficiency.

6. The electric car of Exercise 24 weighs 1100 lb including driver. Estimate the required motor rating in hp and the operating range with a 12-V, 100-A-h battery.
7. Six SCRs are used in the three-phase adjustable voltage motor supply in Fig. 24.22.
   (a) Sketch the waveforms of the supply voltages. (See Fig. 7.23.)
   (b) Under what circumstances will motor current flow through SCRs 1 and 6? 2 and 6?
   (c) Describe the operation of the circuit.
   (d) For a firing angle of 90°, sketch the waveform of the voltage across the motor.
   (e) Repeat part (d) for $\theta = 120°$.

Chapter 25
Alternating Current Machines

Alternators
Synchronous Motors
Induction Motor Operation
Induction Motor Performance

In Chapter 24 we applied our knowledge of electromechanics to practical machines carrying direct currents in the stator and rotor windings. We considered the operating principles of these dc machines and derived simple circuit models for use in predicting their performance under changing conditions.

Now we are going to follow the same approach in studying synchronous machines and induction motors. In the induction motor, alternating current is applied to the stator and alternating currents are induced in the rotor by transformer action. In the synchronous machine, direct current is supplied to the rotor and alternating current flows in the stator. The same machine can function as *alternator* or *motor;* the only difference is in the direction of energy flow.

Alternators

In the alternator of Fig. 25.1, current is transferred to the rotor by stationary brushes riding on sliprings. Because the required field power is much less than that associated with the armature, placing the field on the rotor is the most practical arrangement. In other designs transformer action is used to transfer alternating current to the rotor and there convert it to direct current, using semiconductor diodes mounted on the rotor. This arrangement eliminates the necessity for any sliding contacts and greatly reduces maintenance problems.

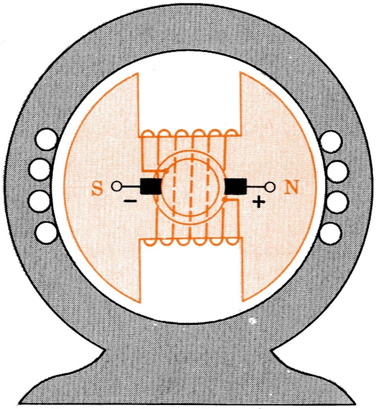


Figure 25.1 An elementary, two-pole synchronous machine.

As pointed out in connection with Fig. 7.23, the emf generated in each turn of the 4-turn coil shown is slightly out of phase with the emf generated in its neighbor because it experiences maximum flux density at a different instant. By proper distribution of the turns and shaping of the field pole the total emf can be made to approach a sinusoid. Two other similar windings (not shown) occupy the remainder of the periphery of the stator and the result is a balanced set of three sinusoids that can be connected in **Y** or **Δ** (see Figs. 7.24 and 7.26).

Basic Relations

In this discussion the emphasis is on three-phase alternators although two-phase machines are sometimes used, particularly in control systems. For convenience we assume that all machines are **Y**-connected and balanced, and we work in terms of *one phase;* the phase voltage is the line-to-neutral voltage ($V_p = V_l/\sqrt{3}$) and the phase torque is just one-third of the total torque. Voltages and currents are sinusoidal and are represented by phasors. Assuming linear magnetic circuits greatly simplifies the mathematics and leads to results that are quite satisfactory for our purposes.

On the basis of these assumptions the basic relations for generated emf and developed torque are:

$$E = kn\Phi_R \qquad \text{and} \qquad \mathrm{T}_d = k_{\mathrm{T}}\Phi_R\Phi_S \sin\delta \tag{25-1}$$

where E and Φ are rms values and T_d is an average value. Flux Φ_R is due to dc field current in the rotor, flux Φ_S is due to ac current flowing in the stator, and δ is the torque angle. In generator operation, the frequency of the emf is directly related to the speed of rotation. A complete cycle of emf is generated as a pair of magnetic poles pass a given coil; therefore, the frequency in hertz is

$$f = \frac{n}{60}\frac{p}{2} \tag{25-2}$$

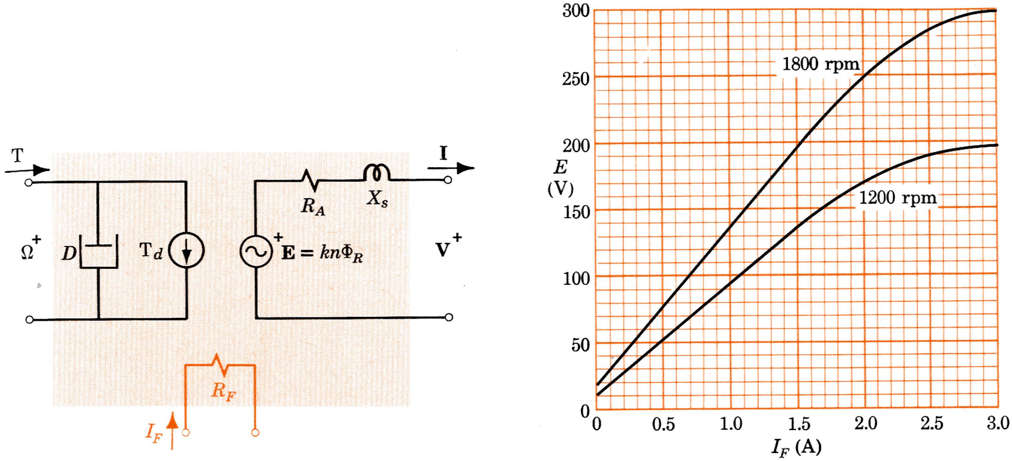
In motor operation, to develop unidirectional torque the rotor must turn at *synchronous speed* n_s given by

$$n_s = \frac{120f}{p} \tag{25-3}$$

It is convenient to remember that for 60-Hz operation the highest synchronous speed is 3600 rpm, obtained with a 2-pole machine ($p = 2$).

Circuit Model of an Alternator

For steady-state operation, the circuit model for the synchronous machine is as shown in Fig. 25.2. (It is quite similar to that for the dc machine in Fig. 24.2.) The phasors **E, I,** and **V** are rms values of per-phase quantities; emf E is defined by a magnetization curve. The *armature resistance* R_A is an effective resistance which represents copper losses and takes into account nonuniform distribution of current in the armature conductors. The *synchronous reactance* X_s represents



(a) Steady-state circuit model

(b) Magnetization curve

Figure 25.2 Representation of an alternator.

the effect of armature reaction and leakage inductance. Armature reaction results in a component of flux that induces a voltage in quadrature with the current and therefore can be represented by an inductive reactance. The effect of stator leakage flux that does not react with the rotor is also represented by an inductive reactance. In our simplified approach, these two effects are combined into a single reactance X_s that is assumed to be constant.†

The assumption regarding X_s holds fairly well for a machine with a uniform air gap. In such a *cylindrical-rotor* machine, the field windings are embedded in slots in an otherwise smooth rotor. The assumption is inaccurate for the more typical *salient-pole* machine shown in Fig. 25.1. In either case, the machine constants can be determined by open- and short-circuit tests just as for the transformer. In all but the smallest machines, the synchronous reactance is much larger than the armature resistance; therefore, we neglect R_A except in calculating losses.

Alternator Characteristics

In considering the electrical characteristics of an alternator driven at constant speed, the simplified model of Fig. 25.3a is satisfactory. Note that R_A is neglected and $E = k'I_F$ for a linear magnetic circuit. The external characteristic (per phase) is defined by

$$\mathbf{V} = \mathbf{E} - \mathbf{I}(R_A + jX_s) \cong \mathbf{E} - j\mathbf{I}X_s \tag{25-4}$$

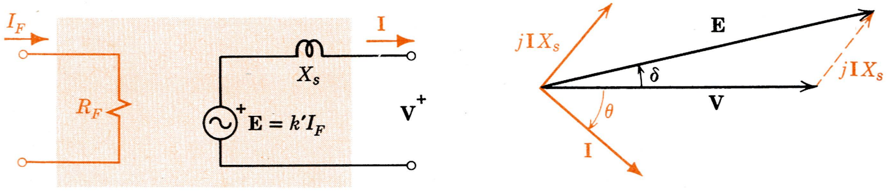
or

$$\mathbf{E} = \mathbf{V} + j\mathbf{I}X_s \tag{25-5}$$

The corresponding phasor diagram for an inductive (assumed) load is shown in

---

†For a more complete discussion of synchronous reactance see Chapter 10 of Robertson and Black, *Electric Circuits and Machines,* 2nd ed., D. Van Nostrand, New York, 1957.


I_F
R_F
X_s
+
$\mathbf{E} = k'I_F$
I
$\mathbf{V}^+$
$j\mathbf{I}X_s$
E
δ
V
θ
I
$j\mathbf{I}X_s$

(a) Circuit model for one phase (b) Phasor diagram

Figure 25.3 Electrical characteristics of an alternator.

Fig. 25.3b. Phasor **V** is taken as a horizontal reference and **I** is drawn at a lagging angle θ. The $j\mathbf{I}X_s$ voltage drop is 90° ahead of **I**, and **E** is the sum of **V** and $j\mathbf{I}X_s$. Since **E** is directly determined by rotor flux Φ_R and **V** bears a similar relation† to air gap flux Φ_S linking the stator, the angle between **E** and **V** is just equal to δ, the angle between Φ_R and Φ_S.

The behavior of the alternator under various power factor conditions is illustrated in Fig. 25.4. We assume that the alternator is supplying a constant power

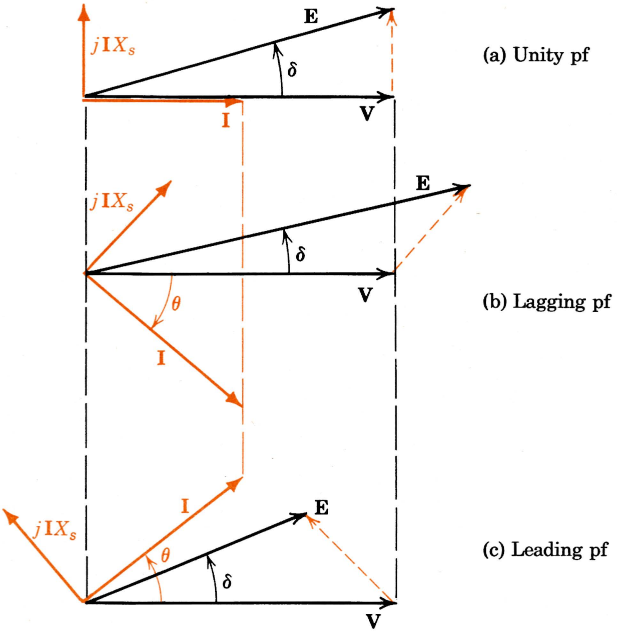

E
$j\mathbf{I}X_s$
δ
I
V
(a) Unity pf
$j\mathbf{I}X_s$
E
δ
V
θ
I
(b) Lagging pf
I
E
$j\mathbf{I}X_s$
θ
δ
V
(c) Leading pf

Figure 25.4 Effect of load power factor on the operation of an alternator supplying constant power $P = VI\cos\theta$.

---

† This statement assumes that the effects of leakage reactance and armature resistance are neglected.

$P = VI\cos\theta$, at a fixed voltage and frequency, as the power factor is changed. At unity pf, E and V are approximately equal in magnitude. For lagging-pf operation, E is considerably larger than V; therefore, I_F must be increased above the value required for a unity-pf load. For leading-pf operation, E is considerably smaller than V and less field current is required. A physical explanation for this behavior is that the leading current in the stator produces a component of mmf (by armature reaction) that aids the mmf established by the field current; therefore, less I_F and E are required for the same V. (Lagging-pf operation is considered in Example 1.)

Example 1

A three-phase 6-pole alternator is rated at 10 kVA, 220 V, at 60 Hz. Synchronous reactance is $X_s = 3\ \Omega$. The no-load terminal voltage (rms) follows the magnetization curve of Fig. 25.2b. Determine the rated speed and predict the field current required for full-load operation at 0.8 lagging pf.

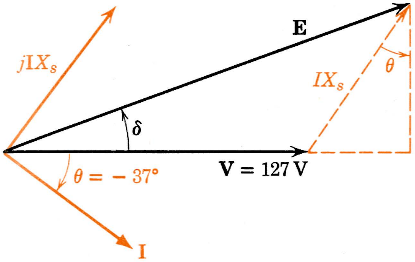


Figure 25.5 Alternator performance.

The rated speed is given by (Eq. 25-3)

$$n_s = \frac{120f}{p} = \frac{120 \times 60}{6} = 1200 \text{ rpm}$$

Assuming **Y** connection, the phase voltage is

$$V_p = \frac{220}{\sqrt{3}} = 127 \text{ V}$$

the rated phase current is

$$I = \frac{P_A/3}{V_p} = \frac{10{,}000/3}{127} = 26.2 \text{ A}$$

and the voltage drop across the synchronous reactance is

$$IX_s = 26.2 \times 3 = 78.6 \text{ V}$$

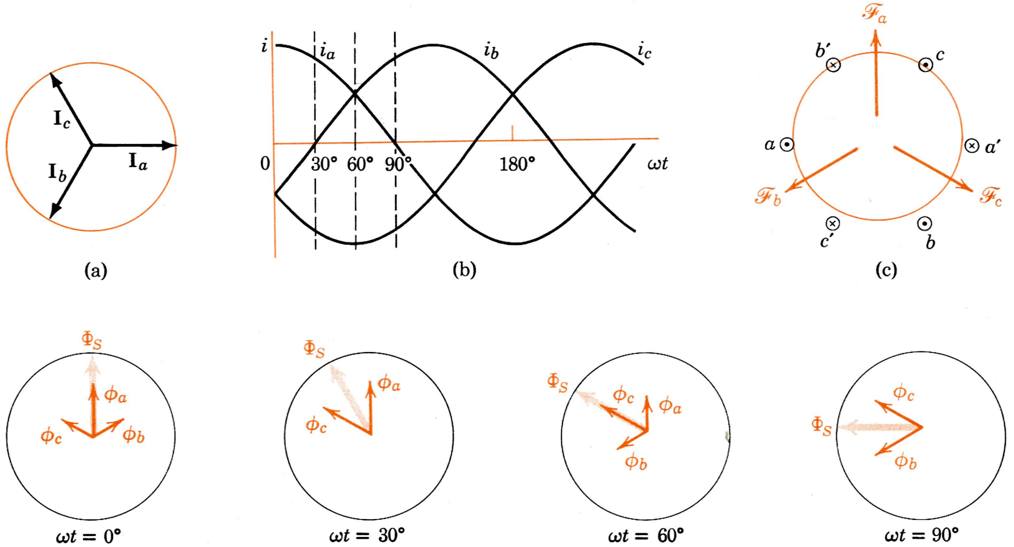
The phasor diagram for 0.8 lagging-pf operation is shown in Fig. 25.5. The necessary value of generated emf is

$$\mathbf{E} = \mathbf{V} + j\mathbf{I}X_s = \mathbf{V} - IX_s\sin\theta + jIX_s\cos\theta$$
$$= 127 + 47.2 + j62.9 \cong 184\,\underline{/20^\circ} \text{ V}$$

An rms voltage of 184 V at 1200 rpm requires 2.3 A.

The Rotating Field

For unidirectional torque, $\sin\delta$ must not change sign (Eq. 25-1) and therefore the stator field must rotate at the same speed as the rotor field. How can a rotating field be produced by stationary coils? This concept is fundamental to the operation of synchronous and induction motors. It can be explained in terms of the mmf contributed by the three balanced currents in Fig. 25.6a. The instantaneous values of the currents are shown in (b). Positive currents in the concentrated



(a) (b) (c)

(d)

Figure 25.6 Production of a rotating stator field.

windings produce mmfs in the directions indicated in (c). At the instant $\omega t = 0°$, say, i_a is a positive maximum and i_b and i_c are negative and one-half maximum. Assuming a linear magnetic circuit, the principle of superposition applies, and the flux contributions of the three currents are as shown in (d). The stator flux Φ_S is the vector sum of the three contributions. At $\omega t = 30°$, the relative magnitudes of the three currents have changed and the position of the resulting stator flux has shifted 30°. We say that a rotating magnetic field exists because the position of maximum flux rotates at synchronous speed.

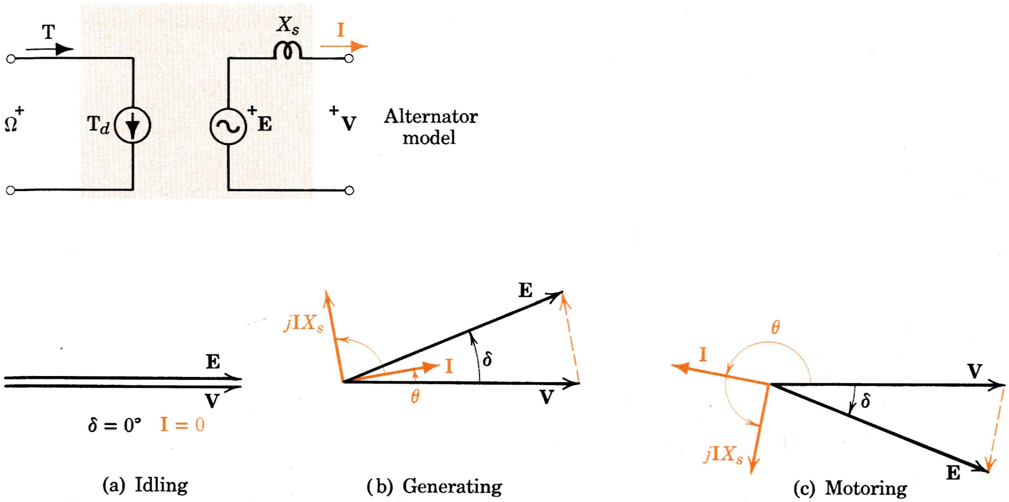
It was pointed out in Chapter 7 that in a balanced three-phase circuit power is constant rather than pulsating, as in a single-phase circuit. We should have expected that an alternator supplying constant electrical power at a constant speed would develop a constant torque. For constant torque, δ must remain constant or the stator flux must move in synchronism with the rotor flux. Figure 25.6 is a demonstration of a situation that we could have anticipated.

Synchronous Motors

Another feature of synchronous machines, which should come as no surprise, is that by changing the direction of power flow the alternator becomes a motor. In a dc machine, the direction of current flow is reversed by changing the field current and thereby changing the generated emf. But Fig. 25.4 indicates that in a synchronous machine there may be large changes in **E** with no change at all in the power delivered by the alternator. How can such a machine be made to deliver torque to a load?

Motor Characteristics

Consider an alternator driven at synchronous speed by a gasoline engine. The output terminals of the alternator are connected to the three-phase line of the local power company. Referring to the model in Fig. 25.7, let **V** be one phase voltage and let I_F be adjusted so that $\mathbf{E} = \mathbf{V}$. Under this condition, $\mathbf{I} = 0$ and there is no power transfer between the alternator and the line; the alternator is idling. Now let us open the throttle of the gasoline engine and attempt to accelerate the rotor. As the rotor is driven ahead of the stator field, there is a voltage difference between **E** and **V**, and current **I** flows to satisfy Eq. 25.5. (Note that the magnitude of **E** does not change.) The electrical power output must be matched by a corresponding mechanical input or



(a) Idling (b) Generating (c) Motoring

Figure 25.7 Power flow in a synchronous machine using the alternator model.

$$P_E = 3VI\cos\theta = P_M = 3\mathrm{T}_d\Omega \qquad (25\text{-}6)$$

where $\mathrm{T}_d = k_\mathrm{T}\Phi_R\Phi_S \sin\delta$ as before. The increase in developed torque associated with angle δ equals the increase in applied torque and the average speed of the engine-alternator combination does not change. (The forward shift in δ accompanying mechanical input is clearly visible when the rotor is illuminated by a stroboscopic light.)

If the engine throttle is returned to the idle position, the torque angle returns to zero and the rotor continues to turn at synchronous speed. If the throttle is closed, the rotor *tends* to slow down; the engine is now, in effect, an air compressor acting as a load on the synchronous machine. As the rotor shifts back through an angle δ† (Fig. 25.7c), the emf **E** is retarded in phase with respect to **V** and a voltage difference exists across X_s. The resulting current **I** is nearly 180° out of phase with **V**. The electrical *output* $VI \cos\theta$ is *negative*, or the machine is absorbing electrical power from the line and delivering mechanical power to the shaft. It is operating as a *synchronous motor.*

---

† Note that δ is in "electrical degrees" where there are 180° electrical between adjacent magnetic poles. In a 4-pole machine, one revolution corresponds to $(p/2) \times 360 = 720$ electrical degrees.

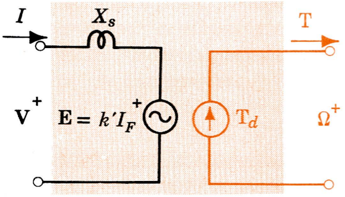


Figure 25.8 Circuit model for a synchronous motor.

A more convenient model for a synchronous motor is shown in Fig. 25.8. Here the reference direction of I has been changed and in-phase current represents positive electrical input. In general, the input power per phase is

$$P = VI \cos\theta \tag{25-7}$$

The voltage relation for a motor is

$$\mathbf{V} = \mathbf{E} + j\mathbf{I}X_s \tag{25-8}$$

and the current is

$$\mathbf{I} = \frac{\mathbf{V} - \mathbf{E}}{jX_s} \tag{25-9}$$

if the armature resistance is neglected. On the basis of this simple model, as interpreted by these equations, the behavior of synchronous motors under various operating conditions can be predicted.

Example 2

The alternator of Example 1 ($X_s = 3\ \Omega$) is rated at 10 kVA, 220 V, and 26.2 A at 1200 rpm. Predict the torque angle and the field current for unity-pf operation as a motor at rated load.

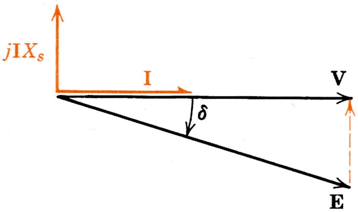


Figure 25.4 Unity-pf operation.

The phasor diagram for unity-pf operation is shown in Fig. 25.9. A slight increase in **E** over the idling value (Fig. 25.7a) is required. Solving Eq. 25-8,

$$\mathbf{E} = \mathbf{V} - j\mathbf{I}X_s = 127 - j26.2 \times 3 \cong 150\underline{/-32^\circ}\ \mathrm{V}$$

and the torque angle δ is 32°.

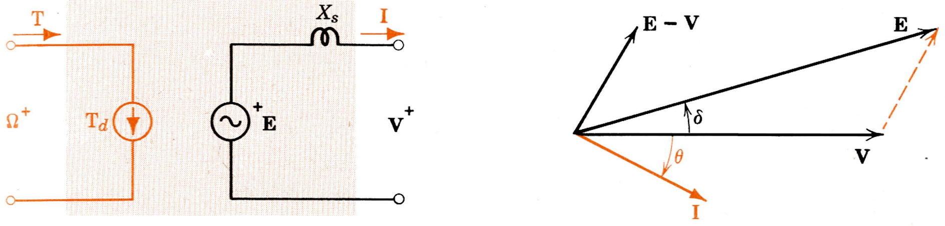
Referring again to the magnetization curve of Fig. 25.2b, we note that a voltage of 150 V at 1200 rpm requires a field current of 1.65 A.

Power Relations

Since torque varies as the sine of the angle δ, there is a maximum possible developed torque and power. If the load on a motor is increased continuously, the rotor field drops further and further behind the stator field until this maximum is reached. Any further increase in load torque pulls the rotor into a position of less than maximum developed torque, and the rotor falls out of step and out of synchronism. At less than synchronous speed the developed torque is first in one direction and then in the other, and the machine comes to a shuddering halt. Some "loads" may overhaul the motor and drive it as an alternator at greater than synchronous speed, pulling it out of synchronism.

What happens if the input to an alternator is increased continuously? An approximate expression for the maximum electrical power developed in an alternator can be derived by using the simplified model and phasor diagram of Fig. 25.10. By Eq. 25-5,

$$\mathbf{I} = \frac{\mathbf{E} - \mathbf{V}}{jX_s} = -j\left[\frac{E\angle\delta}{X_s} - \frac{V\angle 0^\circ}{X_s}\right] = -j\left[\frac{E(\cos\delta + j\sin\delta)}{X_s} - \frac{V}{X_s}\right]$$



(a) Alternator model (b) Phasor diagram

Figure 25.10 Derivation of developed power in an alternator.

The real part of **I** is $(E/X_s)\sin\delta$ which is equal to $I\cos\theta$. Then the electrical power developed per phase is $VI\cos\theta$ or

$$P = \frac{VE}{X_s}\sin\delta \tag{25-10}$$

The electrical power at the air gap is just equal to the developed mechanical power, which is the product of torque and speed. Therefore,

$$T_d = \frac{P}{\Omega} = \frac{VE}{\Omega_s X_s}\sin\delta \tag{25-11}$$

In a motor the electrical power developed is negative, which is consistent with the negative value of δ associated with mechanical power output (Fig. 25.7c).

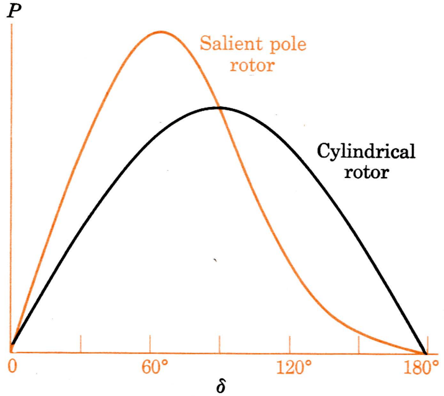


Figure 25.11 Power angle curves for synchronous motors.

The black line in Fig. 25.11 is the power angle curve based on Eq. 25-10 and holds for cylindrical-rotor machines. Most synchronous motors have salient poles and the reluctance torque contribution is appreciable. It can be shown (see Problem 5 in Chapter 23) that the reluctance torque varies as sin 2δ; the colored line in Fig. 25.11 shows the variation expected in salient-pole machines. The advantages of salient-pole construction are a higher maximum or *pull-out* torque and a smaller torque angle for a given load in the normal operating region (below $\delta \cong 30°$).

Example 3

Predict the pull-out torque for the 10-kVA, 220-V, 1200-rpm motor of Example 2 operating at $E = 150$ V.

Assuming cylindrical-rotor construction, Eq. 25-10 applies. For $\delta = 90°$, the maximum total power is

$$P_{\max} = \frac{3VE}{X_s} = \frac{3 \times 127 \times 150}{3} \cong 19{,}000 \text{ W or N-m/s}$$

At 1200 rpm, the pull-out torque is

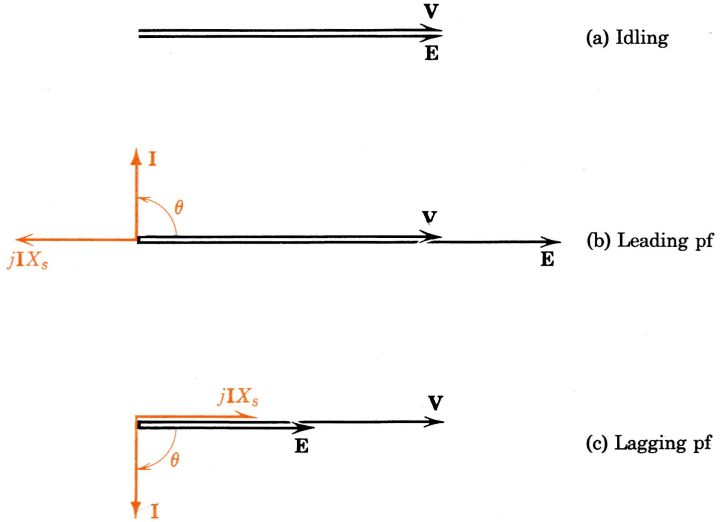
$$T_{\max} = \frac{P_{\max}}{\Omega_s} = \frac{19{,}000 \times 60}{2\pi \times 1200} \cong 150 \text{ N-m}$$

Note that the maximum power is nearly twice the rated value.

Reactive Power Relations

A unique feature of the synchronous motor is its ability to carry shaft load and draw a leading current. In an industrial plant operating at a lagging power factor, use of some relatively expensive synchronous motors may be justified by the savings resulting from improved power factor. (See the discussion preceding Example 3 in Chapter 7.)

The effect of varying field current on the power factor of an unloaded motor is displayed by the phasor diagrams in Fig. 25.12. Because the motor carries no shaft load, the torque angle is zero; for normal excitation $E = V$ and the



(a) Idling

(b) Leading pf

(c) Lagging pf

Figure 25.12 Effect of excitation on the power factor of an unloaded synchronous motor.

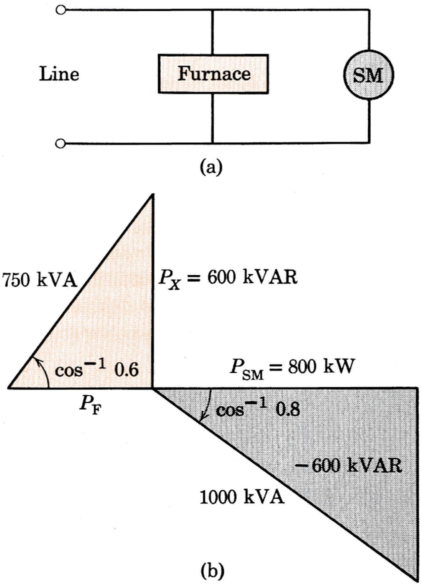
motor draws no current. If, as in Fig. 25.12b, the field current is increased, E exceeds V and the resulting current is determined by Eq. 25-9. Such an *overexcited* machine draws a leading current, i.e., the sinusoidal current reaches its maximum 90° ahead of the sinusoidal terminal voltage. The power drawn is still zero since $VI\cos\theta = 0$, but the reactive power $P_X = VI\sin\theta$ is appreciable. These conditions are characteristic of a capacitor (or condenser) and therefore a rotating machine of this type is called a *synchronous condenser*. (See Example 4 on the following page.)

If the field current is below normal, E is less than V and a lagging current flows (Fig. 25.12c). Such an *underexcited* machine draws a positive reactive power and could be called a *synchronous inductor;* however, underexcited machines do not offer advantages comparable to those of the synchronous condenser. The phasor diagrams in Fig. 25.12 are drawn for a motor with no shaft load; in general, there is shaft load and $\mathbf{I}$ has an in-phase component.

In dc machines, the direction of power flow is determined by the relative magnitude of E and V and can be controlled by I_F. In a synchronous machine, changing the relative magnitude of E affects only the reactive power flow. To change the direction of power flow requires a change in the torque angle δ and this is accomplished by changing the mechanical power supplied to the shaft. Opening the throttle of the driving engine advances the rotor and increases δ, and electrical power is produced. Loading the shaft retards the rotor and increases δ in a negative direction, and mechanical power is produced.

Example 4

A three-phase induction furnace (Fig. 25.13a) draws 750 kVA at 0.6 lagging pf. A 1000-kVA synchronous motor is available. If the overall pf of the combination is to be unity, estimate the mechanical load that can be carried by the motor.



(a)

(b)

Figure 25.13 Power factor correction.

The reactive power of the furnace is

$$P_X = 750 \sin (\cos^{-1} 0.6) = 600 \text{ kVAR}$$

as shown in Fig. 25.13b. By operating at leading pf, the synchronous motor can supply this reactive power. If losses are neglected, the available power of the synchronous motor is

$$P_{SM} = \sqrt{P_A{}^2 - P_X{}^2} = \sqrt{1000^2 - 600^2} = 800 \text{ kW}$$

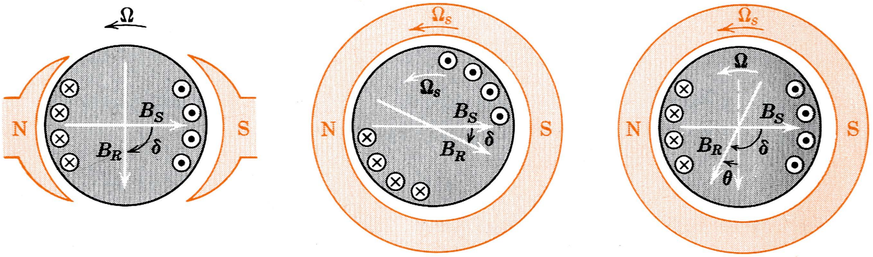
At the cost of a small increase in copper loss in the motor field windings, the necessary reactive power is supplied to the furnace and 80% of the motor rating is available for any mechanical load.

Performance of Synchronous Motors

The synchronous motor is a constant-speed device with a relatively high efficiency and a controllable power factor. A source of dc excitation is required and, since it has no inherent starting torque, special starting provisions are required. Most synchronous motors are started by the induction motor action described in the next section. Commercial designs provide starting torques of the order of 120% of rated and pull-out torques of the order of 175% of rated. One feature is that by increasing the number of poles, low constant speeds are available without gearing. Synchronous motors are ideal for constant-speed applications where high starting torque is not required. They are found in all types of industrial plants, and are used to drive compressors, pumps, crushers, and other loads up to several thousand horsepower. Their range of application is greatly expanded with the adjustable frequency drive described on p. 679.

Induction Motor Operation

In any rotating machine, we can conceive of torque as a result of the interaction of a current and a magnetic field, and we can write $\mathrm{T}_d = k_a B_1 i_2 \sin \delta$ (Eq. 23-27a). In a dc machine, direct current is supplied to the stator to establish B_1 and to the rotor as I_2, and the commutator maintains δ at the optimum 90° at any speed (Fig. 25.14a). In a synchronous machine, polyphase alternating current supplied to the stator produces a rotating magnetic field, and the dc rotor field established by I_2 rotates at synchronous speed to develop a torque proportional to $\sin \delta$ (Fig. 25.14b). In an induction machine, polyphase alternating current supplied to the stator produces a rotating magnetic field (just as in the synchronous machine), and alternating currents are induced in the rotor by transformer action. The interaction between the induced rotor currents and the stator field produces torque.



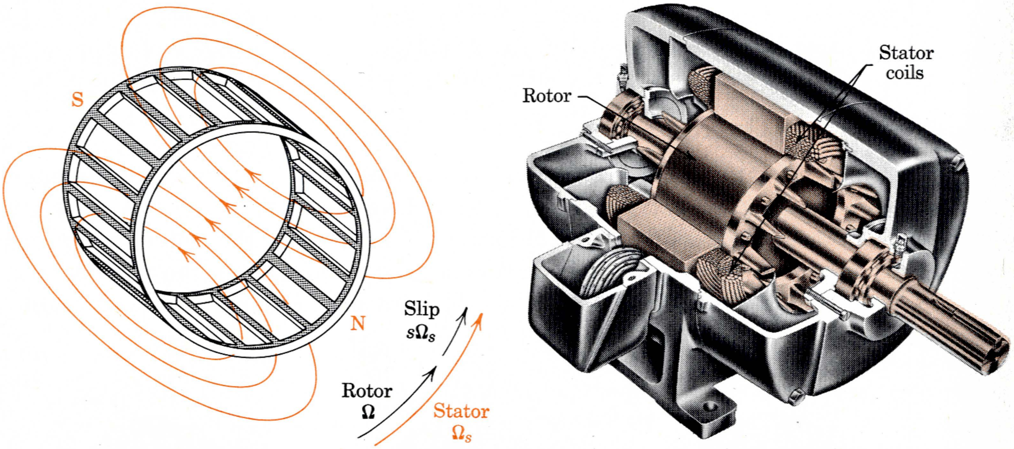
(a) Direct-current (b) Synchronous (c) Induction

Figure 25.14 Torque production in rotating machines.

Because the synchronous machine is much better for generating alternating current, the induction generator is not commercially significant and therefore our emphasis is on the three-phase induction motor. This simple, rugged, reliable, inexpensive device (invented by Nikola Tesla in 1887) consumes the major part of all the electric energy generated. It is built in sizes ranging from a fraction of one horsepower to many thousand horsepower. This important electromagnetic device deserves our attention.

Operating Principles

In its most common form, the rotor consists of axial conductors shorted at the ends by circular connectors to form a *squirrel cage* (Fig. 25.15). The stator is identical with that of a synchronous motor with the windings arranged to provide an even number of poles. As the stator field moves past the rotor



(a) Rotor construction (b) Cutaway view

Figure 25.15 Induction motor with squirrel-cage rotor.

conductors, emfs are induced by transformer action. The induced emf is a maximum in the rotor conductors under the poles, as shown in Fig. 25.14c. Because of the reactance of the rotor windings, the current and therefore the rotor field lag the emf by angle θ, and the torque angle is greater than 90°.

The stator field rotates at synchronous speed Ω_s determined by the frequency and number of poles (Eq. 25-3). The rotor always turns at a lower speed Ω; if the rotor turned at synchronous speed, there would be no change in flux linkage, no induced current, and no torque. The small difference in speed that produces flux cutting and motor action is called the *slip*. By definition, the slip is

$$s = \frac{\Omega_s - \Omega}{\Omega_s} = \frac{n_s - n}{n_s} \qquad (25\text{-}12)$$

and is usually expressed in percent. The slip at full load for a squirrel-cage motor is typically 2 to 5%. Note that the rotor field created by the induced rotor currents moves ahead at a speed $s\Omega_s$ relative to the rotor structure. The absolute speed of the rotor field is

$$\Omega_{RF} = \Omega + s\Omega_s = \Omega_s \qquad (25\text{-}13)$$

as required for unidirectional torque. In other words, the magnetic pole induced in a portion of the rotor is always of the proper polarity to react with the stator pole passing by. The flexibility of the induction motor, in contrast to the fixed-speed character of the synchronous motor, is due to this principle.

Example 5

A three-phase, 220-V, 60-Hz induction motor runs at 1140 rpm. Determine the number of poles, the slip, and the frequency of the rotor current.

Since the synchronous speed must be slightly higher than the running speed and because at 60 Hz the only possibility is 1200 rpm (Eq. 25-3), we conclude that this is a 6-pole machine. At 1140 rpm the slip is

$$s = \frac{n_s - n}{n_s} = \frac{1200 - 1140}{1200} = 0.05 \text{ or } 5\%$$

The induced emf in the rotor is due to the slip speed and therefore

$$f_2 = sf = 0.05 \times 60 = 3 \text{ Hz}$$

Note that at starting $n = 0$ and $s = 1$; therefore, the frequency of the induced starting current is 60 Hz.

Basic Relations

The induction motor is essentially a transformer with a rotating secondary. The force that exists between primary and secondary coils in a transformer appears as useful torque in an induction motor. If the rotor is held stationary (by the inertia of a starting load or by an applied brake), the passing stator field induces in one phase of the rotor winding an emf of magnitude E_2 and frequency f. The current that flows at standstill is

$$\mathbf{I}_2 = \frac{\mathbf{E}_2}{R_2 + j2\pi f L_2} = \frac{\mathbf{E}_2}{R_2 + jX_2} \tag{25-14}$$

where R_2 and L_2 are the effective resistance and inductance of the rotor winding on a per-phase basis.

Now we are going to take advantage of an ingenious method for representing the power-developing capability of a short-circuited coil. At standstill, the slip $s = 1$ and the rotor frequency is just the stator frequency f. At any other speed, the slip is s, the induced emf is sE_2, the rotor frequency is sf, and the rotor reactance is sX_2. In general, the rotor current at slip s can be expressed as

$$\mathbf{I}_2 = \frac{s\mathbf{E}_2}{R_2 + jsX_2} = \frac{\mathbf{E}_2}{R_2/s + jX_2} = \frac{\mathbf{E}_2}{R_2 + jX_2 + R_2[(1-s)/s]} \tag{25-15}$$

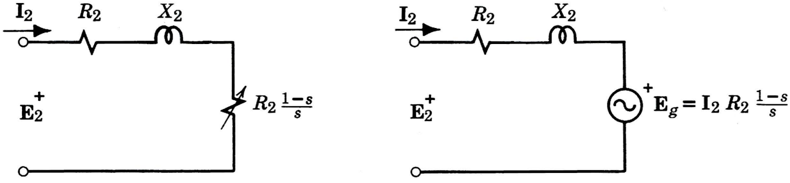
since

$$R_2\frac{1-s}{s} = \frac{R_2}{s} - R_2 \tag{25-16}$$

Then Eq. 25-15 can be rewritten as

$$\mathbf{E}_2 = \mathbf{I}_2(R_2 + jX_2) + \mathbf{I}_2 R_2 \frac{1-s}{s} = \mathbf{I}_2\mathbf{Z}_2 + \mathbf{E}_g \tag{25-17}$$

Two circuit models satisfying Eq. 25-17 are shown in Fig. 25.16. The first looks like the secondary of a transformer with a variable load resistance; the second represents the load as a reversible energy transformation. In each model, the electrical power transferred to the rotor is $E_2 I_2 \cos\theta_2$. The power lost irreversibly is $I_2^2 R_2$; the remainder, $E_g I_2$, is converted into mechanical power.



(a) Controlled load resistance (b) Controlled voltage source

Figure 25.16 Equivalent circuit models of the rotor of an induction motor.

In analyzing the energy conversion taking place in an induction motor, the basic relations are

$$E_g = I_2 R_2 \frac{1-s}{s} \qquad \text{and} \qquad \mathrm{T}_d = \frac{E_g I_2}{\Omega} \tag{25-18}$$

The first of these relations results from a consideration of electrical energy in the rotor; the second results from the application of the energy-conservation principle to the electromechanical conversion process. To obtain an expression for torque in the form of Eq. 23-27a, consider the total electrical power transferred from the stator at Ω_s to the rotor. From either model of Fig. 25.16 this is $P_2 = E_2 I_2 \cos\theta_2$. But E_2 is the rms value of an induced emf directly proportional to B_1, and $\cos\theta_2$ is just equal to $\sin\delta$ (Fig. 25.14c); therefore,

$$\mathrm{T}_2 = \frac{P_2}{\Omega_s} = \frac{1}{\Omega_s} E_2 I_2 \cos\theta_2 = k_a B_1 I_2 \sin\delta \tag{25-19}$$

The developed torque T_d is less than T_2 by an amount corresponding to the $I_2{}^2 R_2$ power loss.

Deriving a Circuit Model

The model for the rotor (Fig. 25.16) resembles that of the secondary of a transformer. If we use the complete transformer model of Fig. 22.19a, the circuit model of an induction motor is as shown in Fig. 25.17. Since R_2 and X_2 are important only insofar as they affect input electrical power, it is convenient to

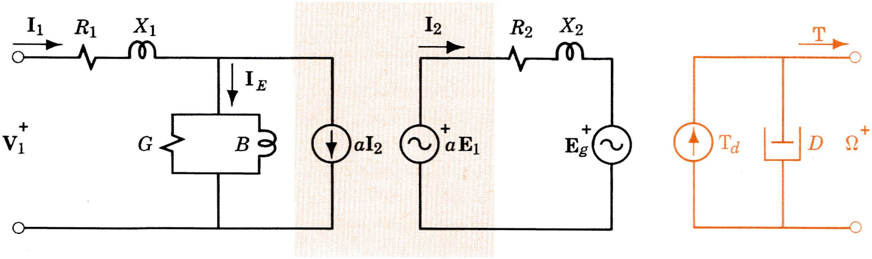
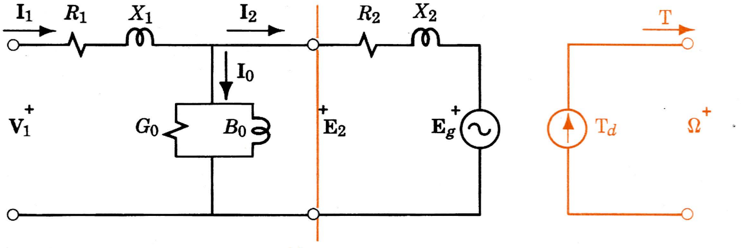


Figure 25.17 Step 1: A general circuit model for an induction motor.

assume that the turn ratio a is unity and to define R_2 and X_2 on that basis. If this is done, the ideal transformer consisting of aI_2 and aE_1 is superfluous and can be omitted. The colored line in Fig. 25.18 corresponds to the air gap across which energy is coupled to the rotor. A further simplification results if we transfer the rotational losses $\Omega^2 D$ to the primary and combine them with the stator core loss represented by G. (The rotor core loss is usually negligibly small because of the low rotor frequency.) The sum of rotational loss and core loss is conveniently determined by measuring the no-load input where I_0 is the no-load current.



Air gap

Figure 25.18 Step 2: A circuit model based on no-load measurements.

One other modification reduces the labor of performance calculations and simplifies the general analysis without introducing excessive error. Just as for the transformer (see Fig. 22.18), moving the elements representing no-load loss to the input terminals results in a circuit consisting of two branches in parallel across a fixed voltage. In the transformer the reluctance of the magnetic circuit is small and the exciting current is only a small percentage of rated current. In an induction motor, however, the air gap raises the reluctance and the rotational loss adds to the core loss. As a result, the no-load current is 20 to 40% of rated current, and this is too large to neglect. The effect of the no-load current can be taken into account by adjusting the terminal voltage to account for the $I_0(R_1 + jX_1)$ drop as indicated in the following example.

Example 6

The parameters of a 25-hp, 220-V, 60-Hz induction motor are: $R_1 = 0.05\ \Omega$, $R_2 = 0.06\ \Omega$, $X_1 = 0.23\ \Omega$, $X_2 = 0.23\ \Omega$. The no-load power input is 690 W at 220 V and 20 A. Calculate the adjusted voltage V_A in Fig. 25.19 and determine the error made by using an algebraic rather than a phasor calculation.

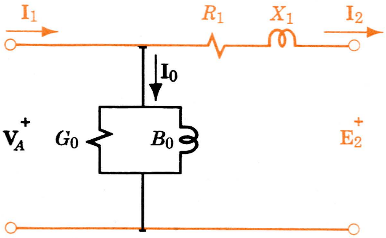


Figure 25.19 Step 3: Using an adjusted terminal voltage.

The stator impedance is

$$R_1 + jX_1 = 0.05 + j0.23 \cong 0.25\angle 78^\circ\ \Omega$$

At no load

$$P_0 = 3V_1 I_0 \cos\theta_0$$

or

$$\theta_0 = \cos^{-1}(690/3 \times 127 \times 20) = -85^\circ$$

and

$$\begin{aligned} \mathbf{V}_A &= \mathbf{V}_1 - \mathbf{I}_0(R_1 + jX_1) \\ &= 127\angle 0^\circ - 20\angle -85^\circ\,(0.25\angle 78^\circ) \\ &= 127\angle 0^\circ - 5\angle -7^\circ = 127 - 4.9 + j0.6 \\ &= 122.1 + j0.6 \cong 122.1\ \text{V} \end{aligned}$$

Algebraically,

$$\begin{aligned} V_A &= V_1 - I_0\sqrt{R_1^2 + X_1^2} \qquad (25\text{-}20) \\ &= 127 - 20\sqrt{(0.05)^2 + (0.23)^2} \\ &= 127 - 20(0.25) = 122\ \text{V} \end{aligned}$$

The error is very small.

As indicated by Example 6, the error made in using an algebraic calculation of V_A is negligibly small and therefore Eq. 25.20 is satisfactory. (Draw the phasor diagram approximately to scale and you will see why.) In many cases R_1 can be neglected compared to X_1 in this calculation, and in some cases use of the unadjusted terminal voltage is satisfactory.

One final simplification is desirable. Combining Eqs. 25-12 and 25-18, we can write

$$\mathrm{T}_d = \frac{E_g I_2}{\Omega} = \frac{I_2^2 R_2[(1-s)/s]}{\Omega_s(1-s)} = \frac{I_2^2 R_2}{s\Omega_s} \qquad (25\text{-}21)$$

Since $E_g = I_2 R_2(1-s)/s$, the total series resistance is

$$R_1 + R_2 + R_2\frac{1-s}{s} = R_1 + \frac{R_2}{s} \qquad (25\text{-}22)$$

The reactances X_1 and X_2 are difficult to separate and usually they are assumed equal; the sum is then $X = 2X_1$. The result is the simplified circuit model of Fig. 25.20.

Induction Motor Performance

On the basis of the simplified model we can predict the performance of induction motors under varying conditions.

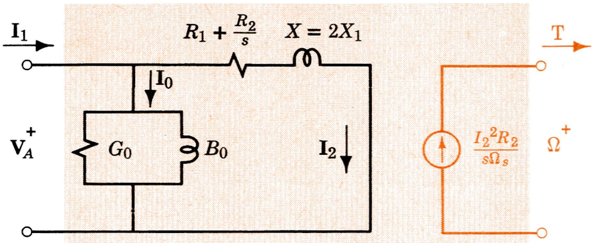


Figure 25.20 Final step: A simplified circuit model for an induction motor.

Speed-Torque Characteristics

An important characteristic of any motor is the speed-torque relation. For the induction motor, the total torque T is $3T_d$ or

$$\mathrm{T} = \frac{3I_2{}^2R_2}{s\Omega_s} = \frac{3V_A{}^2}{\Omega_s}\frac{R_2/s}{(R_1 + R_2/s)^2 + X^2} \tag{25-23}$$

The speed-torque curve for a typical squirrel-cage motor is plotted in Fig. 25.21. Under normal operating conditions the slip is small (less than 5%) and the induction motor is essentially a constant-speed machine under load. For s small, R_2/s is large compared to R_1 and to X, and a good approximation is

$$\mathrm{T}_{\text{normal}} \cong \frac{3V_A{}^2s}{\Omega_sR_2} \tag{25-24}$$

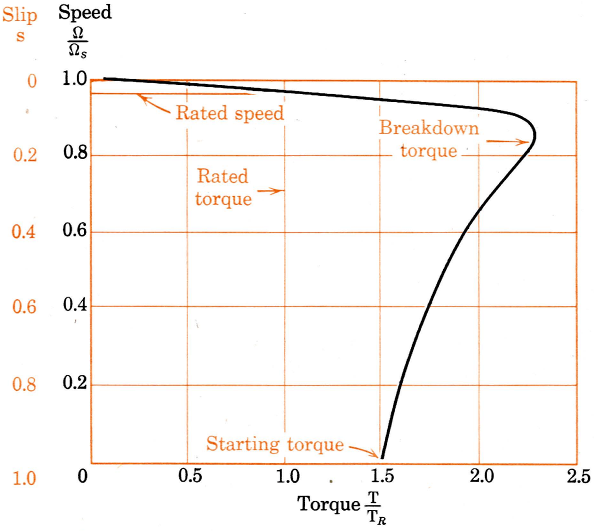


Figure 25.21 Typical speed-torque curve for a squirrel-cage induction motor.

In the normal operating range, torque is directly proportional to slip. Another conclusion from Eq. 25-24 is that at normal speeds the torque is inversely proportional to R_2. The use of larger rotor conductors permits larger induced currents and greater torque is developed at a given slip.

When the speed is low, $s \cong 1$ and the reactive term in Eq. 25-23 is large compared to the resistive term. An approximation for low-speed torque is

$$\mathrm{T}_{\text{low speed}} \cong \frac{3V_A{}^2R_2}{s\Omega_s X^2} \tag{25-25}$$

and we see that near standstill, torque varies inversely with slip. At standstill, $s = 1$ and Eq. 25-25 becomes

$$\mathrm{T}_{\text{starting}} \cong \frac{3V_A{}^2R_2}{\Omega_s X^2} \tag{25-26}$$

We conclude that starting torque is directly proportional to $V_A{}^2$ and to R_2.

Maximum torque occurs at an intermediate speed; how could this maximum be predicted? If we differentiate Eq. 25-23 with respect to slip and set the derivative equal to zero, the slip for maximum torque is found to be

$$s = \frac{R_2}{\sqrt{R_1{}^2 + X^2}} \tag{25-27}$$

At this value of slip, the maximum or *breakdown* torque is

$$\mathrm{T}_{\max} = \frac{3V_A{}^2}{2\Omega_s(R_1 + \sqrt{R_1{}^2 + X^2})} \tag{25-28}$$

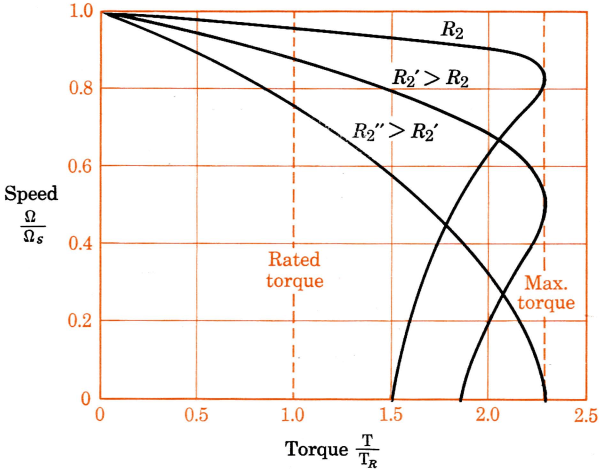


Figure 25.22 The effect of R_2 on speed-torque curves.

An analysis of the effect of R_2 leads to the following conclusions:

1. At normal speeds, torque is inversely proportional to R_2.
2. At low speeds, torque increases with increasing R_2.
3. The slip for maximum torque is directly proportional to R_2.
4. The maximum torque is independent of R_2.

These conclusions are supported by the speed-torque curves for similar motors with varying values of R_2 shown in Fig. 25.22. Apparently the design of the squirrel-cage rotor is a compromise. Low rotor resistance results in low slip for a given torque and high operating efficiency; high rotor resistance results in high starting torque and relatively low starting current. (Why?)

Example 7

Predict the starting torque and maximum torque for the 6-pole motor of Example 6 ($R_1 = 0.05\ \Omega, R_2 = 0.06\ \Omega, X = 0.46\ \Omega$, and $V_A = 122$ V) and estimate the rotor resistance for maximum torque at starting.

By Eq. 25-26, the starting torque is

$$T_{\text{start}} \cong \frac{3V_A^2 R_2}{\Omega_s X^2} = \frac{3(122)^2 0.06}{(1200 \times 2\pi/60)(0.46)^2} = 101 \text{ N-m}$$

By Eq. 25-28, the maximum torque is

$$T_{\max} = \frac{3V_A^2}{2\Omega_s(R_1 + \sqrt{R_1^2 + X^2})}$$

$$\cong \frac{3(122)^2}{2 \times 40\pi(0.05 + 0.46)} \cong 360 \text{ N-m}$$

since R_1^2 is small compared to X_2^2.

To obtain maximum torque at starting, R_2 is given by Eq. 25-27 for $s = 1$ or

$$R_2 = \sqrt{R_1^2 + X^2} = \sqrt{(0.05)^2 + (0.46)^2} \cong 0.46\ \Omega$$

Motors with Variable Rotor Resistance

How can the insertion of additional resistance in the rotor increase the starting torque? At first thought it would seem that insertion of resistance would decrease the induced current and therefore the torque. But at starting, $s = 1$ and $R_2/s = R_2$, which is small compared to X_2, so the magnitude of the current (Eq. 25-15) is affected only slightly by an increase in R_2. In contrast, the power factor ($\cos\theta$ in Eq. 25-19) is greatly improved and the torque is increased. An alternative viewpoint is that δ is brought closer to 90° and the current I_2 (and therefore B_2; see Fig. 25.14c) bears a more favorable relation to the flux density B_1.

How can the resistance be increased during starting for high torque and reduced to a minimum during normal operation for low slip and high efficiency?

One method is to arrange connections to the rotor coils (by using slip rings and brushes) so that additional resistances can be connected in series. In such a *wound-rotor* motor, the rotor winding is similar to that on the stator; an auxiliary *controller* permits the insertion of variable external resistances corresponding to R_2' and R_2'' in Fig. 25.22. The wound-rotor motor provides optimum starting conditions (high torque and low current) and also provides a measure of speed control, since for a given torque the speed depends on rotor resistance. The disadvantages of the wound-rotor motor include higher first cost, higher maintenance cost, and lower efficiency during variable-speed operation.

The *double-squirrel-cage* motor is a clever design for obtaining some of the advantages of variable resistance without the disadvantages. Each rotor conductor consists of two parts, one in the normal location and a second deeply imbedded in the rotor iron (Fig. 25.23). The leakage reactance and therefore the impedance of the inner conductor are higher than those of the outer conductor. At starting, rotor frequency is high and little current flows in the inner conductor; the effective rotor resistance is high. At normal speeds, slip is low, rotor frequency is low, the two conductors are in parallel, and the effective rotor resistance is low. (How would the speed-torque curve of such a motor differ from those in Fig. 25.22?)

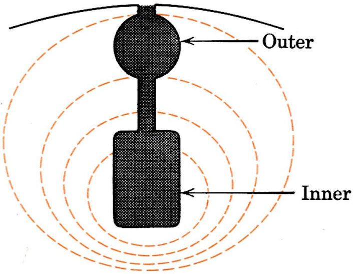


Figure 25.23 Double-rotor conductors and leakage flux.

Performance of Induction Motors

The similarity between the induction motor and the transformer has been noted; in fact, a wound-rotor motor can be used as a static transformer with a three-phase secondary voltage available at the slip rings. Another similarity is in the method of determining the machine parameters in the circuit model. A *no-load* motor test provides data for the determination of G_0 and B_0 just as an open-circuit transformer test. Corresponding to the short-circuit transformer test is the *blocked-rotor* test; with the rotor held stationary, all the input power is dissipated as losses and R and X can be determined.

The stator reactance X_1 is usually assumed to be half of X, but the division of R must be done carefully because R_2 is a key factor in motor performance. One approach is to measure the dc resistance between terminals which gives an approximate value of $2R_1$, assuming a **Y**-connected stator. The dc value is then corrected empirically to give an effective ac value of R_1 and the remainder of R is the effective value of R_2 referred to the stator circuit (see Problem 8). When the model parameters are known, the performance of any motor may be predicted from the mathematical relations or from the circuit model itself.

Example 8

Predict the input, output, and efficiency of the 1200-rpm motor in Example 6 at a slip of 2.8%. For the model in Fig. 25.24 the parameters are: $V_A = 122$ V, $R_1 = 0.05\ \Omega, R_2 = 0.06\ \Omega$, and $X = 0.46\ \Omega$. The no-load power input is 690 W at 20 A and 220 V.

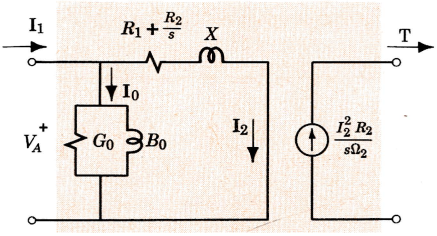


Figure 25.24 Induction motor performance calculations.

At $s = 2.8\%$, the rotor current is

$$\mathbf{I}_2 = \frac{V_A}{R_1 + R_2/s + jX} = \frac{122\underline{/0^\circ}}{0.05 + 0.06/0.028 + j0.46}$$

$$= \frac{122\underline{/0^\circ}}{2.19 + j0.46} \cong 54.4\underline{/-12^\circ}\ \text{A}$$

and

$$\mathbf{I}_1 = \mathbf{I}_0 + \mathbf{I}_2 = 20\underline{/-85^\circ} + 54.4\underline{/-12^\circ}$$

$$= 63.1\underline{/-29.5^\circ}\ \text{A}$$

The total input power is

$$3V_1I_1 \cos\theta = 3 \times 127 \times 63.1 \times 0.87 = 20.9\ \text{kW}$$

The total developed torque is (by Eq. 25-23)

$$\mathrm{T} = \frac{3I_2{}^2R_2}{s\Omega_s} = \frac{3(54.4)^2 \times 0.06}{0.028 \times 40\pi} \cong 152\ \text{N-m}$$

and the output power is

$$P_{\text{hp}} = \frac{\mathrm{T}\Omega}{746} = \frac{152 \times (1 - 0.028)40\pi}{746} \cong 25\ \text{hp}$$

The efficiency is

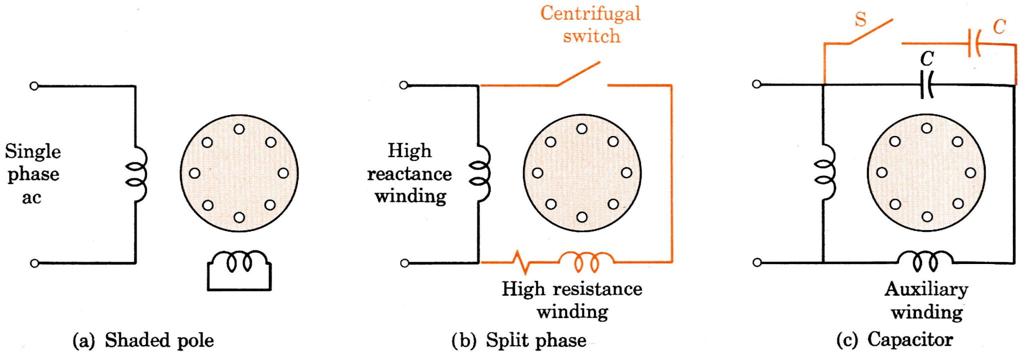
$$\frac{\text{Output}}{\text{Input}} = \frac{25 \times 0.746}{20.9} = 0.89 \text{ or } 89\%$$

The polyphase squirrel-cage induction motor is a rugged, efficient, easily started, nearly constant-speed machine that is applicable to many purposes. The wound-rotor version provides control of speed and starting torque in return for an increased investment. Induction motors are particularly useful for hoists, mixers, or conveyors that require a smooth start under load. An interesting variation with possible applications for the future is the *linear induction motor* in which the stator windings are arranged on a flat surface to give a linearly moving field.

Single-Phase Induction Motors

To obtain the rotating field necessary for unidirectional torque, a polyphase winding is required. With a single-phase winding, an oscillating field is possible, but the axis of the stator field remains fixed in space. If a squirrel-cage rotor is turning in such an oscillating field, a pulsating torque is developed, which tends to bring the rotor to nearly synchronous speed. If the rotor is stopped and then started in the opposite direction, the developed torque is in the opposite direction and again the rotor is accelerated. The problem in single-phase motor design is to get the rotor started.

There are several ingenious methods for doing this.† In the *shaded-pole* motor (Fig. 25.25a), a heavy copper coil is wound around half of each salient stator pole. Induced currents in the shorted turn delay the buildup of magnetic flux in that region of the pole. The magnetic flux vector appears to shift as a function of time and the rotor experiences the effect of a partially rotating field.



(a) Shaded pole (b) Split phase (c) Capacitor

Figure 25.25 Single-phase induction motors.

The *split-phase* motor employs two separate windings having different reactance-resistance ratios. The current reaches its maximum in the high-reactance winding at a later time and the rotor experiences a shift in magnetic field which provides the necessary starting torque. When the motor is nearly up to speed, the high-resistance winding is disconnected by a centrifugal switch.

The *capacitor* motor employs a capacitor in series with an auxiliary winding to provide the necessary phase shift. For improved performance two capacitors are used. The larger provides good starting torque and then is switched out by a centrifugal switch. The smaller remains in the circuit to provide good operating efficiency and power factor.

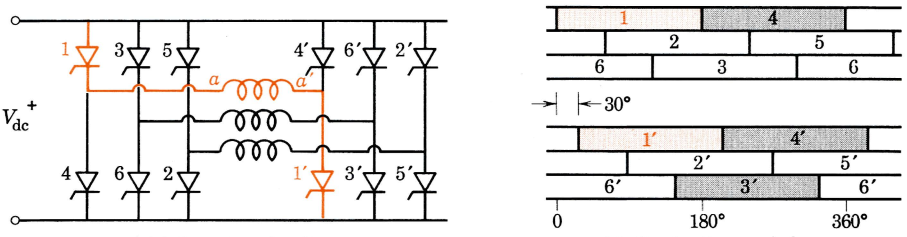
Adjustable Speed Drives

The application of ac motors, synchronous or induction, can be greatly expanded by providing an adjustable frequency power supply. The SCR makes it possible to provide stepless speed control at high efficiency and reasonable cost so that the flexibility of dc machines can be obtained without their inherent maintenance problems and speed limitations.

---

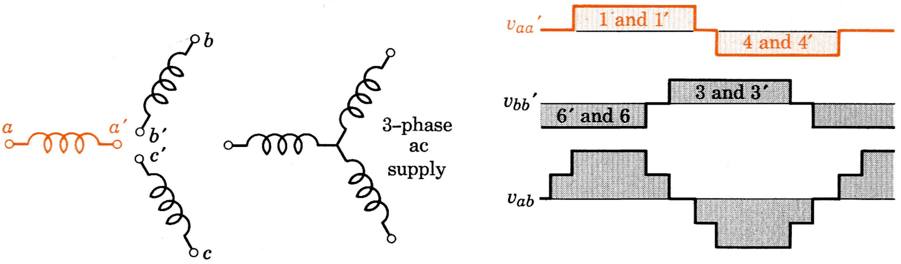
† For further description see Chapter 9 of Skilling, *Electromechanics,* John Wiley & Sons, New York, 1962, or Chapter 12 of Robertson and Black, *Electric Circuits and Machines,* 2nd ed., D. Van Nostrand, New York, 1957.

One approach is to use an SCR *rectifier* to convert ac power to direct current and an SCR *inverter* to provide three-phase, adjustable frequency, adjustable voltage power. In the inverter shown in Fig. 25.26a, six pairs of controlled rectifiers are used to connect the primaries of a transformer to the dc supply. If SCRs 1 and 1′ are fired so that they are ON at the same time, the dc voltage is applied to winding *a–a′*; when SCRs 4′ and 4 are ON, the dc voltage is applied to the same winding with the reverse polarity. The transformer input is an alternating square wave whose frequency is determined solely by the firing or *gate circuit.*



(a) Inverter circuit

(c) Conduction periods



(b) Transformer connection

(d) Voltage waveforms

Figure 25.26 Adjustable frequency supply using an SCR inverter.

The magnitude of output voltage is reduced by retarding the conduction period of the second set of SCRs. In Fig. 25.26c, the delay angle is 30° and the simultaneous conduction angle is 150°. The resulting waveforms of phase voltages v_{aa} and v_{bb}, and the corresponding line-to-line voltage, are shown in Fig. 25.26d. The transformer output voltage will approximate a sine wave. To assure adequate torque at all speeds (and maintain optimum flux densities), the output voltage is increased as the frequency increases.

With an adjustable frequency drive, speed regulation is precise, starting is smooth, and acceleration or deceleration may be precisely timed. Several motors may be run in synchronism at any selected speed. These characteristics are particularly desirable in the steel and textile industries.

SUMMARY

- In a synchronous machine, the basic relations are:

  $$E = kn\Phi = k'I_F \qquad \text{and} \qquad \mathrm{T}_d = k_\mathrm{T}\Phi_R\Phi_S \sin\delta$$

  The speed of the rotating stator field is $n_s = 120f/p$.
  Unidirectional torque is developed only when the rotor turns at n_s.
  The direction of power flow depends on torque angle δ.
  Power transformations are explained by phasor diagrams.

- In an alternator, $\mathbf{E} = \mathbf{V} + j\mathbf{I}X_s$ and δ is positive.
  The power factor of the load determines the required E and I_F.
  Mechanical power is absorbed when the rotor is advanced.

  $$P_M = 3\mathrm{T}_d\Omega = P_E = 3VI\cos\theta = 3\frac{VE}{X_s}\sin\delta$$

- In a synchronous motor, $\mathbf{V} = \mathbf{E} + j\mathbf{I}X_s$ and δ is negative.
  Motor power factor is determined by E and therefore by I_F.
  Electrical power is absorbed when the rotor is retarded.
  Pull-out torque occurs at $\delta = 90°$ with cylindrical rotor.
  Has no starting torque; usually employs induction motor action to start.

- In an induction motor, the slip of the rotor coils behind the rotating stator field induces rotor currents that develop torque; $s = (n_s - n)/n_s$.
  The basic energy conversion relations are:

  $$E_g = I_2R_2\frac{1-s}{s} \qquad \text{and} \qquad \mathrm{T}_d = \frac{E_gI_2}{\Omega} = \frac{I_2{}^2R_2}{s\Omega_s}$$

- Use of an adjusted voltage V_A simplifies the induction motor circuit model. In terms of model parameters the total three-phase torque is

  $$\mathrm{T} = \frac{3V_A{}^2}{\Omega_s}\frac{R_2/s}{(R_1 + R_2/s)^2 + X^2}$$

  Useful approximations for normal operation and for starting are:

  $$\mathrm{T}_{\text{normal}} \cong \frac{3V_A{}^2s}{\Omega_sR_2} \qquad \text{and} \qquad \mathrm{T}_{\text{starting}} \cong \frac{3V_A{}^2R_2}{\Omega_sX^2}$$

  Maximum torque is independent of R_2, but it occurs at a speed determined by R_2. In the wound-rotor motor, R_2 is adjustable.

- Single-phase induction motors depend on a partial rotation of the stator field for starting torque; the running torque is pulsating.

REVIEW QUESTIONS

1. Distinguish between a salient-pole and a cylindrical-rotor machine.
2. Explain the nature of synchronous reactance.
3. Why is the angle between **E** and **V** equal to that between Φ_R and Φ_S?
4. What happens to alternator voltage when the load pf changes?
5. Explain the production of a rotating field by stationary coils.
6. How does a synchronous motor react to an increase in shaft load?
7. What happens if the input to an alternator is increased continuously?
8. How does reluctance torque affect the pullout torque of a motor?
9. How can a synchronous motor be used to correct power factor?
10. How would you reverse the rotation of a synchronous motor? An induction motor? A dc motor?
11. Discuss and compare the value of δ in commutator, synchronous, and induction motors.
12. Explain how a short-circuited rotor produces torque.
13. Specify several possible operating speeds for 50-Hz motors.
14. Why is the rotor core loss usually negligibly small?
15. Draw the phasor diagram for Example 6 to scale.
16. Why is induction-motor torque directly proportional to slip at some speeds and inversely proportional at other speeds?
17. Outline an experimental procedure for determining induction-motor parameters.
18. Examine and classify a single-phase motor and explain its operation.
19. Explain how an inverter changes dc to polyphase ac.
20. How is the frequency adjusted in an inverter?

EXERCISES

1. Determine the operating speed of the following 3-phase alternators:
   (a) 12,000 V, 24 pole, 60 Hz.
   (b) 440 V, 4 pole, 50 Hz.
   (c) 220 V, 6 pole, 400 Hz.
2. Determine the number of poles of the following 3-phase alternators:
   (a) 600 V, 600 rpm, 60 Hz.
   (b) 440 V, 750 rpm, 50 Hz.
   (c) 200 V, 12,000 rpm, 400 Hz.
3. A portable, single-phase ac generator is rated at 5 kW at 125 V at 0.8 pf lagging ($R_A = 0.15\,\Omega$, $X_s = 0.6\,\Omega$). Represent the generator by a model and draw phasor diagrams for rated current at: (a) lagging, (b) unity, and (c) leading pf.
4. At a hydroelectric power plant, the 3-phase 600-rpm alternators are rated at 360,000 kW at 18,000 V line-to-line and 0.85 pf lagging. The synchronous reactance per phase† is 1 Ω. For rated conditions (60 Hz), determine:
   (a) The number of poles and the current.
   (b) The generated emf and power angle.
   (c) The phasor diagram.
5. A 3-phase, 6-pole synchronous machine is rated at 10 kVA, 220 V, 60 Hz. The magnetization curve (volts rms) is similar to Fig. 25.2b and X_s per phase is 4 Ω.†
   (a) Draw an appropriate circuit model for alternator operation.
   (b) Estimate the field current for rated voltage at no-load.
   (c) Repeat part (b) for full load at 0.8 pf leading.
   (d) Repeat part (b) for full load at 0.8 pf lagging.
6. A 40-kVA, 220-V, 3-phase alternator requires a field current of 2 A for rated voltage at no load. With $I_F = 2$ A and $\mathbf{I}_A = 100\underline{/-70^\circ}$ A, the terminal voltage is only 133 V. Draw the phasor diagram and evaluate X_s.

---

† All 3-phase machines are assumed to be **Y**-connected.

7. Continue Fig. 25.6d by drawing the flux vectors corresponding to $\omega t = 180°$ and $210°$. On the stator of Fig. 25.6c, show the location of N and S poles for $\omega t = 180°$.
8. (a) Given a three-phase line with current phasors as shown in Fig. 25.6a, construct a wiring diagram similar to that shown in Fig. 25.6c so that a three-phase synchronous motor will turn in a clockwise direction.
   (b) Assume only two of the phase windings of the motor are connected to the line in part (a). Investigate the possibility of producing a circular rotating stator field by adjusting the relative amplitudes of the two phase voltages.
9. The machine of Exercise 5 is operating as a synchronous motor at rated voltage and current. The field current is increased slightly and it is noticed that the stator current decreases. Draw an appropriate phasor diagram and determine whether the motor is operating with unity, lagging, or leading power factor. Explain your reasoning.
10. In deriving a circuit model for predicting the behavior of a synchronous motor, how is the model affected by: (a) neglecting electrical losses, (b) neglecting mechanical losses, (c) assuming a linear magnetization curve, (d) assuming steady-state operation. Draw a circuit model incorporating the above simplifications.
11. A synchronous motor is represented by a simple circuit model. With the motor unloaded, the field current is adjusted until the stator input current I is a minimum ($E \cong V$). Then the motor is loaded until the torque angle δ is $20°$.
    (a) Draw a labeled phasor diagram for the electrical end of the model, using input voltage as a horizontal reference.
    (b) If the field current is then decreased by 30%, draw the new phasor diagram and predict the new power angle and power factor.
12. A synchronous motor carries a load that results in a power angle $\delta = 30°$; field current is adjusted for minimum stator current.
    (a) Draw a labeled model and phasor diagram for this condition.
    (b) If the load torque is halved, predict the effect on speed, power angle, power factor, and line current.
13. Blackouts may occur if heavily loaded alternators are suddenly disconnected from the line to prevent their being pulled out of synchronism. Predict the pull-out power of the alternator of Exercise 4.
14. For the alternator of Exercise 5:
    (a) Draw a simplified model and calculate the rated line current.
    (b) Draw a phasor diagram and estimate the generated emf and the torque angle under rated conditions.
    (c) Estimate the pull-out torque for this machine.
    (d) List all assumptions made in the previous calculations.
15. When operating under rated conditions at unity power factor, a synchronous motor draws 30 A from the ac line; the field current is 2 A and $\delta = 20°$. For a constant load (i.e., power output), calculate two additional points and sketch a curve of I versus I_F.
16. The per-phase ratings of a synchronous motor are: 300 V, 15 A, and $X_s = 10\ \Omega$. Field current for minimum stator current with no shaft load is 2 A. For operation as a synchronous condenser, estimate the field current required for full-rated kVAR. Draw a labeled phasor diagram.
17. A plant load consisting of 850 kW at 0.6 lagging pf is supplied by a 1600-kVA transformer. An additional 400 kW of power is required. If the additional load is carried by synchronous motors at 0.8 pf leading, will a new transformer be required? What will be the new plant pf?
18. Determine the number of poles, the slip, and the frequency of the rotor currents at rated load for 3-phase, 10-hp induction motors rated at:
    (a) 220 V, 60 Hz, 1728 rpm.
    (b) 220 V, 50 Hz, 900 rpm.
    (c) 120 V, 400 Hz, 5700 rpm.
19. For a 3-phase, 60-Hz, 15-hp induction motor at rated speed of 1740 rpm, determine:
    (a) The speed of the rotor with respect to the stator (rpm).
    (b) The speed of the rotor field with respect to the rotor.
    (c) The speed of the rotor field with respect to the stator field.

(d) The frequency of the currents in the rotor.

20. The rotor of the motor in Exercise 19 is 20 cm in diameter and 15 cm long. The maximum air-gap flux density is 1 T. Stating any necessary assumptions, estimate the emf (rms) induced in each rotor conductor at standstill and at full load.

21. The no-load data of Example 6 are used with Fig. 25.18 to calculate the adjusted voltage V_A in the simplified model of Fig. 25.19.

(a) Draw, to a large scale, a phasor diagram with V_1 as a horizontal reference showing I_0, I_0Z_1, and V_A.

(b) Explain, briefly, why the algebraic calculation of V_A is nearly as accurate as the vector calculation.

(c) How does use of the model of Fig. 25.19 simplify calculations under load conditions?

22. The per-phase parameters for a 6-pole, 60-Hz, 220-V induction motor are: $R_1 = R_2 = 0.1\,\Omega$ and $X_1 = X_2 = 0.5\,\Omega$. The no-load current is $\mathbf{I}_0 = 17\underline{/-78^\circ}$ A. Predict the input current and the horsepower output at 1140 rpm.

23. A 100-hp, 3-phase, 440-V, 4-pole, 60-Hz induction motor has the following per-phase parameters: $R_1 = 0.06\,\Omega$, $R_2 = 0.08\,\Omega$, $X_1 = X_2 = 0.3\,\Omega$. The no-load power input is 3420 W at 45 A.

(a) Draw a simplified circuit model and label it.

(b) Calculate B_0 and G_0.

(c) Predict the line current, the developed torque, and the horsepower output at 2% slip.

24. For the motor of Exercise 23, predict:

(a) The starting torque.

(b) The breakdown torque.

25. (a) Estimate the speed of the motor in Exercise 18a at 6 hp output.

(b) Repeat for the motor in Exercise 18c at 12 hp output.

26. Estimate the starting torque and the breakdown torque of the induction motor of Exercise 22.

27. Two squirrel-cage induction motors are identical except that the rotor of motor A is made of aluminum of conductivity σ and the rotor of motor C is made of copper of conductivity $1.5\ \sigma$. The starting torque of motor A is 120 N-m. Predict the starting torque of motor C.

28. A 3-phase induction motor (440 V, 60 Hz, 6-pole) has the following per-phase parameters: $R_1 = 0.3\,\Omega$, $X_1 = 1\,\Omega$, $R_2 = 0.2\,\Omega$, $X_2 = 1\,\Omega$. $G_0 = 20$ m℧, and $B_0 = -60$ m℧.

(a) Estimate the no-load input current and power.

(b) Predict the power output at 5% slip.

(c) Predict the motor power factor at 5% slip.

29. For the induction motor in Exercise 28 estimate:

(a) The power output at 3% slip.

(b) The no-load speed, assuming that half the no-load input is due to rotational losses.

30. For the induction motor in Exercise 28 estimate:

(a) The starting torque.

(b) The starting torque if the line voltage drops to 400 V.

31. A 100-hp, 440-V, 1200-rpm, 60-Hz synchronous motor is operated from an adjustable frequency supply. For operation at 330 V and 45 Hz with the same stator current and field current, draw the phasor diagrams and estimate the speed, torque, and power output.

32. Specify the type of electric motor that will meet the following requirements best:

(a) Provide constant speed under all loads.

(b) Provide an easily adjusted speed which remains nearly constant (within 10%) over a wide range of loads.

(c) Operate from ordinary "house current" and be cheap. (Starting torque is not important.)

(d) Operate from a dc line and provide relatively high starting torque.

(e) Provide rugged, maintenance-free operation and nearly constant speed (within 5%) over a wide range of loads.

(f) Operate from the single-phase ac line and provide relatively high starting torque.

(g) Operate on ac, provide high starting torque, and permit some variation of speed under load.

(h) Draw a leading pf current while driving a mechanical load.

PROBLEMS

1. The inductive element in Fig. 25.27 might be a transmission line or the air gap of a rotating machine. If $\mathbf{E}_1 = E_1\underline{/0^\circ}$ and $\mathbf{E}_2 = E_2\underline{/\alpha}$ where E_1 and E_2 are constant and α is a variable angle, derive a general expression for the power transferred and determine α for maximum power transfer.

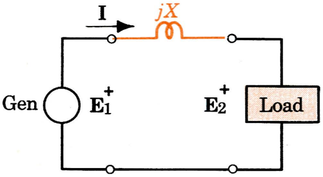


Figure 25.27 Power transfer.

2. Predict the dynamic behavior of a synchronous motor with rotor inertia J turning at speed Ω_s when a sudden load ΔT is applied to the shaft.
   (a) Write a simplified expression for T_d as a function of δ for constant V and Φ_R and small values of δ.
   (b) Write an expression for Ω during the time δ is changing to accommodate an increased load.
   (c) Neglecting frictional resistance to changes in δ, derive an equation governing the dynamic behavior $\delta(t)$ under a suddenly applied torque.
   (d) Determine the form of the variation of δ with time after the torque is applied and draw a conclusion.
3. A synchronous motor is delivering 1000 hp to a constant load; I_F is adjusted for minimum stator current and $\delta = 30^\circ$. Explaining your reasoning with a large-scale phasor diagram, predict the effect of a 20% increase in I_F on (a) speed, (b) power angle, (c) power factor, (d) line current.
4. A 3-phase synchronous motor is rated at 100 hp, 440 V, 131 A, 1200 rpm at 0.8 pf leading; $X_s = 2\ \Omega$ and $I_F = 6$ A for rated conditions.
   (a) What reactive power can be supplied while the motor carries full mechanical load?
   (b) If the motor is operated as a synchronous condenser without shaft load, estimate the field current required and the capacitive kVAR available.
5. A transmission line can be modeled (Fig. 25.28 on a per-phase basis) by an inductive reactance $X_T = 20\ \Omega$. The load consists of a resistance $R_L = 100\ \Omega$. For a terminal voltage $V = 3000$ V, the generator voltage is $V_G = 3060$ V.

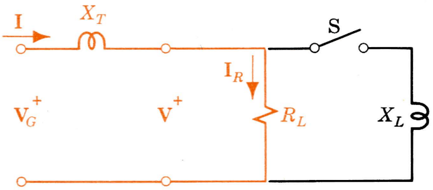


Figure 25.28

   (a) Draw a phasor diagram for switch S open.
   (b) When an inductive load $X_L = 75\ \Omega$ is connected in parallel with R_L, the terminal voltage drops to approximately 2480 V. Keeping the same V phasor as a reference, modify the phasor diagram to show why under the new condition V_G must be considerably larger than V.
   (c) A synchronous condenser ($X_s = 25\ \Omega$), whose magnetization curve is described by $E = 200I_F$, is available for connecting across the terminals at V so that the terminal voltage will be 3000 V with X_L connected. Modify the phasor diagram to include the effect of the synchronous condenser and specify the necessary field current I_F.
   (d) Discuss the use of synchronous condensers in voltage regulation.
6. Determine just enough points to sketch curves showing speed, line current, power factor, and efficiency versus power output for the motor of Exercise 28 in the normal operating region.
7. Laboratory test of a three-phase, **Y**-connected, 6-pole, 150-hp, 4000-V induction motor gave the following results:
   No-load: 4000 V, 7 A, 7 kW
   Blocked rotor: 1340 V, 32 A, 18.2 kW
   Effective ac resistance between terminals = 1.4 Ω
   (a) Determine the model parameters:
   (b) Predict the power output at 1152 rpm.
   (c) Predict the efficiency and power factor at rated load.

Chapter 26

Automatic Control Systems

Control System Characteristics
Dynamic Responses
Feedback Control Systems

One characteristic of life in a technologically developed nation is our use of great amounts of power. A second distinguishing characteristic is our ability to control that power precisely. At one extreme of the power spectrum is the multimegawatt electric power system so precisely controlled that a simple reluctance motor driven by the system becomes an accurate timing device. At the other extreme are the millimicrowatt impulses sent by a cardiac pacemaker to stimulate a defective heart into rhythmic, life-saving contractions. Between these extremes are innumerable applications of control to heating systems, assembly plants, metallurgical processes, and oil refineries and to traffic flow, water supply, medical treatment, aircraft landing, and satellite guidance.

Despite their great variety, control systems can be analyzed into a few basically similar components. In Watt's *flyball governor,* perhaps the first *automatic* control

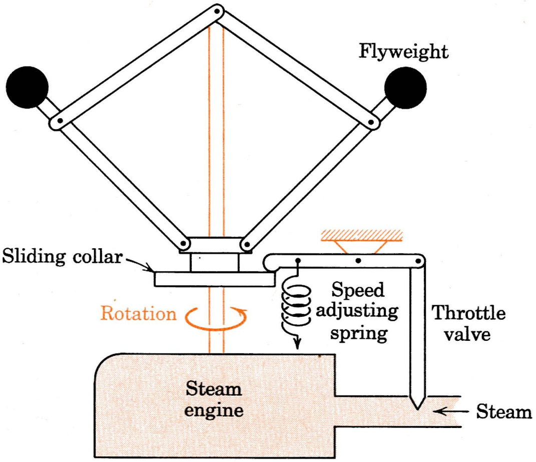


Figure 26.1 A flyball governor.

device, a rotating spindle carrying flyweights (Fig. 26.1) is driven by the governed engine. The engine speeds up until the centrifugal force of the flyweights overcomes the force of the speed-adjusting spring and partially closes the throttle valve. An increase in load on the engine momentarily reduces its speed, reduces the centrifugal force of the flyweights, allows the throttle valve to open, and accelerates the engine until the set speed is reached. The components of the governor are shown in the block diagram of Fig. 26.2. General terms for the functions performed and the variables existing at various points in the system are shown in italics.

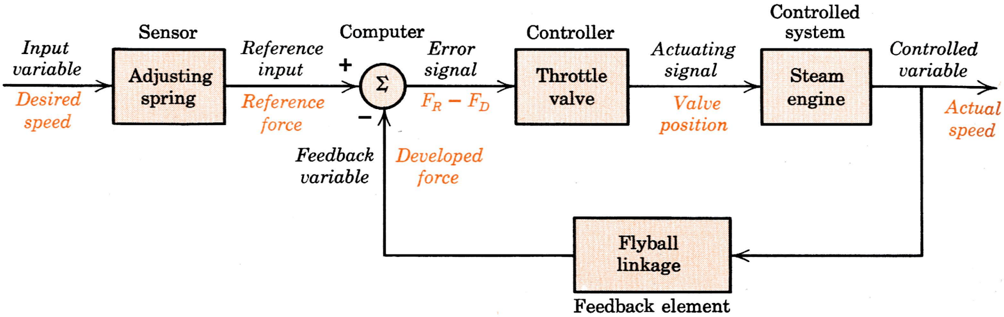

Input variable
Desired speed
Sensor
Adjusting spring
Reference input
Reference force
Computer
+
Σ
−
Error signal
$F_R - F_D$
Feedback variable
Developed force
Controller
Throttle valve
Actuating signal
Valve position
Controlled system
Steam engine
Controlled variable
Actual speed
Flyball linkage
Feedback element

Figure 26.2 Block diagram of a flyball governor.

Most of the elements employed in control systems are familiar to us from our previous study. Transducers are used to measure variables and convert them to electrical form. Computers or comparators are used in determining the difference between reference and feedback variables. Controllers and amplifiers are used to develop signals to actuate the controlled system. Motors are used to provide the "muscle" for system control. Feedback elements are used to facilitate the comparison of output and reference variables. Note that only signals appear in the block diagram of a control system. Power inputs (steam to the engine of Fig. 26.2 or heat from the furnace in Fig. 26.3) are essential to the operation, but ordinarily they do not enter into control calculations.

The study of automatic control represents a continuation of our analysis of linear systems, the central topic of this book, and it provides an opportunity to apply our knowledge of circuit principles, device characteristics, and system techniques. First, we look at the operation of some illustrative examples to discover the essential characteristics of control systems, the benefits obtained by using feedback, and the problems that arise in system design. Then we extend our methods of circuit analysis to permit handling systems incorporating a variety of different elements, each described by a transfer function. Using the new method, we next determine the dynamic behavior of some electromechanical devices. Finally, we consider how to obtain control systems with the necessary accuracy and stability.

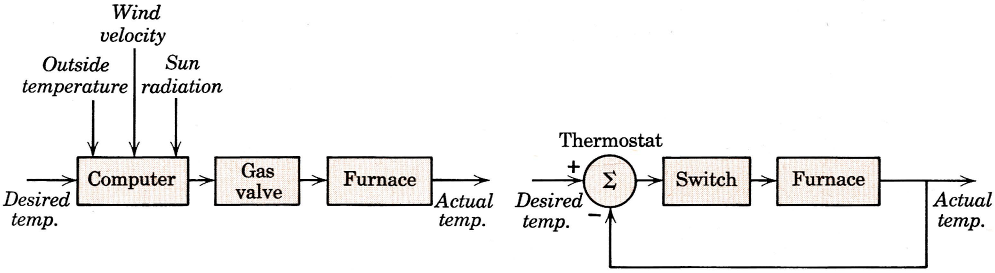
Control System Characteristics

The flyball governor is an example of a *regulator* that maintains a speed, frequency, or voltage constant within specified limits. In a regulator the reference is constant, although it may be adjustable. In a *follower* such as the "self-balancing potentiometer" (see Fig. 27.20), the variation of the output duplicates the variation of the input. If the output variable is a translational or angular displacement such as the aircraft heading in an autopilot (Fig. 9.1), the follower is called a *servomechanism.* Regulating and following can be performed accurately and rapidly using systems employing feedback.

Open-Loop and Closed-Loop Systems

An ammeter is an example of a useful *open-loop* device. To obtain the desired accuracy, the magnet is carefully formed and patiently aged, the moving coil is exactly designed, and the mechanism is precisely fabricated. After the instrument is calibrated it is assumed that the output deflection is an accurate indication of the input current.

A similar device could be used to control the heating of a home by providing a fuel valve opening directly related to the outside temperature. For comfort, however, the valve controller should also take into account such factors as wind velocity, sun radiation, and the number of people in the room. If sufficient data were available, some of these factors could be taken into account by a computer (Fig. 26.3a), but we suspect that there must be a better method of control. In the *closed-loop* system of Fig. 26.3b, a simple thermostat is used to compare the actual temperature with the desired temperature. Whenever the actual is less than the desired temperature, the furnace is ON; otherwise it is OFF.† This simple on-off control system illustrates two general benefits of closed-loop systems. It



(a) Open-loop (b) Closed-loop

Figure 26.3 Temperature control systems.

---

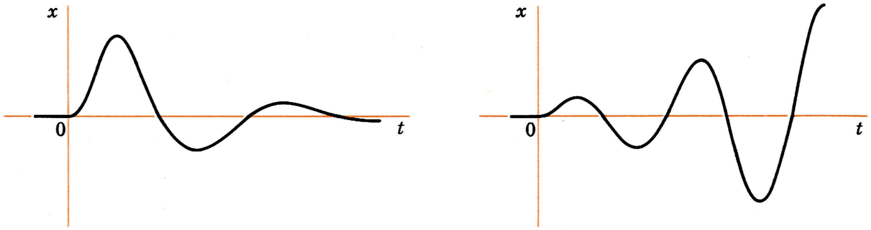
† In practice there is some *dead zone* in which the temperature difference is not sufficient to activate the thermostat.

automatically takes into account factors whose effects are known imprecisely or not at all, and it provides accurate overall performance using inexpensive elements of low individual accuracy.

The Control Problem

A characteristic of closed-loop systems which may be disadvantageous is revealed if we assume that the home is heated by means of hot water circulating in pipes buried in a concrete slab. The energy storage capacity of the slab introduces a time lag in the response of the heating system. If the time lag is several hours, as it might well be, the slab would not begin to radiate heat until long after the thermostat turned ON, and it would continue to radiate long after the thermostat turned OFF. The combination of the amplification needed for accurate control and the energy storage inherent in any physical system may result in erratic behavior.

Providing a feedback path in a system containing amplification may create instability. In a *stable* system the response to an impulse disturbance dies away as time increases (Fig. 26.4a). In an *unstable* system a sudden disturbance may give rise to sustained oscillations or to an uncontrolled increase in a critical variable. By definition, an unstable system is incapable of control. Unfortunately, the requirements of accuracy and stability are incompatible and this creates a difficult design problem for the control engineer.



(a) Stable **(b) Unstable**

Figure 26.4 Impulse response of different systems.

For high accuracy, only a small error signal (see Fig. 26.2) can be tolerated (in the steady state) and therefore the controller must possess high amplification. With high amplification, however, the correction is great and this may lead to instability. Stability can be restored by auxiliary elements that increase the cost and weight and reduce the reliability. The control engineer's problem is to provide the necessary accuracy and speed of response with adequate stability and reliability at a minimum cost and weight.

System Response

One approach to design is to synthesize a possible system, analyze its performance, and compare its performance with the specifications. The performance in terms of accuracy and stability may be deduced from the response of the system to various inputs. In predicting the response, it is convenient to assume that the system is linear (i.e., describable by a linear differential equation) and that the input is an impulse, a step, or a sinusoid.

The basic control system with negative feedback is shown in Fig. 26.5 with sinusoidal signals represented by phasors and transfer functions shown as complex quantities. The controlled variable **C** is fed back through feedback element **H** and the signal **B** = **HC** compared to the reference input **R.** In control system nomenclature, the ratio of return signal **B** to error signal **E** is the *open-loop transfer function* **GH.** For this form of the basic feedback circuit, the *closed-loop gain* is

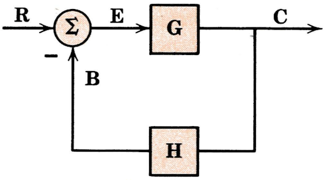


Figure 26-5 The basic feedback control system.

$$\mathbf{G}_F = \frac{\mathbf{G}}{1 + \mathbf{GH}} \tag{26-1}$$

A great deal can be learned about a system from its frequency response, i.e., the amplitude and phase of the output signal in response to an input signal of constant amplitude and variable frequency. From our study of feedback we know that if at any frequency the open-loop gain **GH** has a magnitude of unity and a phase angle of ±180°, oscillation buildup is possible. The *Nyquist stability criterion* is based on this principle and establishes the necessary conditions for stability.

The Nyquist criterion is applied to steady-state forced response data. An alternative approach is to consider the natural response of the system as revealed by a pole-zero plot in the complex plane. We know from our study of circuits that the poles of the admittance function of a one-port correspond to natural response current components. In Chapter 9 we extended this concept to include transfer functions of the two-ports employed in control systems. By plotting the possible locations of the poles of the transfer function as system parameters are varied, the so-called *root locus,* we can determine whether a system is stable, and if it is not, what changes can provide stability.

Dynamic Response of Electromechanical Devices

Control systems are useful only insofar as they are able to cope with changing conditions. In evaluating a regulator, we are interested in its *dynamic* behavior, its response to a sudden change in load, for example. In a follower there may be a change in the input variable or there may be a *disturbance* at some point in the system. In general, such changes are random and therefore unpredictable except on a statistical basis. For our purposes, a good approach is to determine the dynamic behavior described by the complete response to a step function. An alternative approach is to consider the frequency response for steady-state sinusoidal inputs. In either approach we rely on the transfer function.

Transfer Functions

The transfer function of a two-port is the ratio of two signals measured at the output and input ports and described by exponentials of the form $a(t) = Ae^{st}$. The two signals may be entirely different in physical nature, but the transfer function is derived and interpreted just as an admittance or impedance function. Block diagram representation is convenient.

Since transfer functions are derived from linear ordinary differential equations, devices characterized by transfer functions must be represented by models made up of lumped, linear elements. The circuit models we have used to represent electrical and mechanical devices satisfy this requirement, so transfer functions can be determined directly as in Example 1.

In some cases further simplification is possible by ignoring factors that have negligible effect on the behavior being investigated. In Example 1, the inductance of the meter coil was ignored for this reason. Once the transfer function is obtained, the complete step response can be determined following the procedure outlined in Chapter 6. The forced response is calculated using the transfer function evaluated at $s = 0$. (How is $s = 0$ related to the step function?) The natural response consists of terms identified with poles of the transfer function. (Why not use the zeros?) The undetermined coefficients are evaluated from initial conditions. In a control system the initial conditions are usually zero, corresponding to a state of equilibrium prior to the initiation of the change.

First-Order Systems

The order of a system is defined by the highest power of s in the denominator of the transfer function. Frequently encountered first-order transfer functions and their step responses (see Eqs. 9-17 and 9-19) are:

$$T(s) = \frac{ks}{s + \alpha} \qquad y(t) = Ye^{-\alpha t} \tag{26-2}$$

$$T(s) = \frac{k}{s + \alpha} \qquad y(t) = Y(1 - e^{-\alpha t}) \tag{26-3}$$

Example 2 indicates that a shunt motor with constant field current behaves as a first-order system in responding to a step armature voltage input. (See Example 2 on page 692.)

Second-Order Systems

In the transfer function of the d'Arsonval movement in Example 1, s appears to the second power in the denominator. By definition, this is a second-order

Example 1

(a) Write the transfer functions for each element of the circuit model of a d'Arsonval movement shown in Fig. 26.6a.

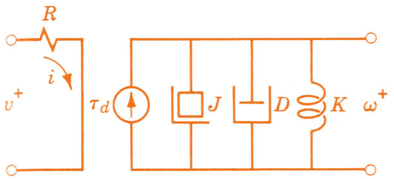
The governing equations and the corresponding transfer functions are:

$$v = Ri \qquad \frac{I(s)}{V(s)} = \frac{1}{R}$$

$$\tau_d = k_T i \qquad \frac{\mathrm{T}_d(s)}{I(s)} = k_T$$

$$\theta = \int \omega \, dt \qquad \frac{\Theta(s)}{\Omega(s)} = \frac{1}{s}$$

$$\tau_d = J\frac{d\omega}{dt} + D\omega + \frac{1}{K}\int \omega dt \qquad \frac{\Omega(s)}{\mathrm{T}_d(s)} = \frac{1}{sJ + D + 1/sK}$$



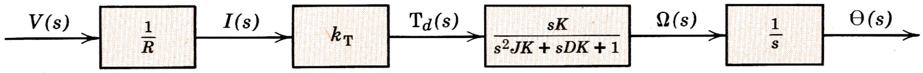
(a) Circuit model

(b) Draw a block diagram for this device.

If the last transfer function is rearranged,

$$\frac{\Omega(s)}{\mathrm{T}_d(s)} = \frac{1}{sJ + D + 1/sK} = \frac{sK}{s^2JK + sDK + 1}$$

and the block diagram is as shown in Fig. 26.6b.



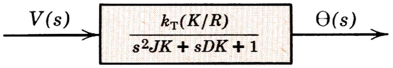
(b) Block diagram

(c) Derive the transfer function for the relation between deflection and applied voltage.

The transfer functions here correspond to the block "gains" employed in Chapter 9. Following the reasoning displayed in Fig. 9.4, we conclude that the overall transfer function of blocks in cascade is the product of the individual transfer functions or

$$T(s) = \frac{\Theta(s)}{V(s)} = \frac{1}{R} \times k_T \times \frac{sK}{s^2JK + sDK + 1} \times \frac{1}{s}$$

$$= \frac{k_T(K/R)}{s^2JK + sDK + 1}$$



(c) Transfer function

Figure 26.6 Deriving the transfer function of a d'Arsonval movement.

(d) Write the transfer function for sinusoidal excitation.

For sinusoidal excitation,

$$\mathbf{T} = \mathbf{T}(j\omega) = \frac{\mathbf{\Theta}}{\mathbf{V}} = \frac{k_T(K/R)}{1 - \omega^2JK + j\omega DK}$$

Example 2

A dc shunt motor with constant field excitation is used to position an indicator in a servomechanism. The following simplifying assumptions are reasonable:

Inductance of armature circuit is negligibly small.

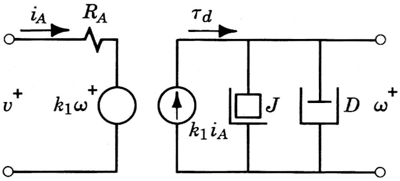
Magnetic flux is constant and $K\phi$ is replaced by k_1.

Total friction of motor and load is represented by D.

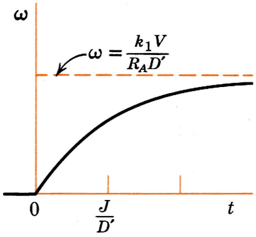
Total inertia of motor and load is represented by J.

Load torque is negligible compared to D and J.

Derive the transfer function and predict the dynamic speed response.



(a) Dynamic circuit model



(b) Step response

Figure 26.7 Dynamic response of a first-order system.

On the basis of these assumptions, the circuit model of Fig. 24.9 is modified as shown in Fig. 26.7a. The governing differential equations are

$$J\frac{d\omega}{dt} + D\omega = \tau_d = k_1 i_A \qquad (26\text{-}4)$$

and

$$i_A = \frac{v - k_1\omega}{R_A}$$

Solving for v and collecting terms, we obtain

$$v = \frac{R_A}{k_1}\left(J\frac{d\omega}{dt} + D\omega + \frac{k_1^2}{R_A}\omega\right)$$

Noting that k_1^2/R_A has the same dimensions as the friction coefficient D, let us define an *equivalent friction coefficient* D' where

$$D' = D + \frac{k_1^2}{R_A}$$

Making this substitution and assuming exponential variation of v and ω, we can write the transfer function as

$$T(s) = \frac{\Omega(s)}{V(s)} = \frac{k_1/R_A}{sJ + D'} \qquad (26\text{-}5)$$

This equation is of the same form as Eq. 26-3. The forced response to a step of magnitude V is (for $s = 0$)

$$\omega_f = \frac{k_1 V}{R_A D'}$$

The denominator has a pole at $s = -D'/J$. Therefore, the total step response is

$$\omega = \frac{K_1 V}{R_A D'}(1 - e^{-D't/J}) \qquad (26\text{-}6)$$

as shown in Fig. 26.7b. The speed ω increases until it is limited by equivalent friction torque alone.

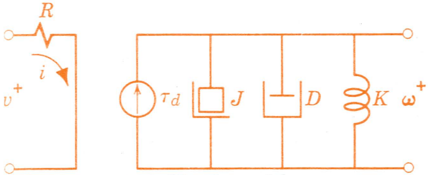
system; the possibility of energy storage in two different forms suggests a more complicated response. However, the transfer function approach is still convenient as illustrated in Example 3.

Normalized Response of a Second-Order System

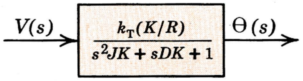
The response anticipated in Example 3 is encountered so frequently in control system analysis that it deserves detailed attention. As before (Eq. 4-18b), we define

Example 3

The circuit model and the transfer function of the d'Arsonval movement of Example 1 are shown in Fig. 26.8. With the movement at rest, a dc voltage V is suddenly applied at $t = 0$. Predict the step response of the movement.



(a) Circuit model



(b) Transfer function

Figure 26.8 Dynamic response of a d'Arsonval meter.

For dc, $s = 0$ in the transfer function and the forced response is

$$\Theta_f = T(0) \times V = \frac{k_T K}{R} V$$

To find the poles of the transfer function, we find the roots of the denominator. Where

$$s^2 JK + sDK + 1 = 0 \tag{26-7}$$

the roots are

$$s_1, s_2 = -\frac{D}{2J} \pm \sqrt{\left(\frac{D}{2J}\right)^2 - \frac{1}{JK}} \tag{26-8}$$

In general, the complete response is

$$\Theta = \frac{k_T K}{R} V + A_1 e^{s_1 t} + A_2 e^{s_2 t} \tag{26-9}$$

Reasoning by analogy from the mathematically similar electrical circuit, we predict that the form of the response is dependent upon the nature of the roots s_1 and s_2. If the roots are real and unequal, the response is over-damped (Fig. 4.8). If the roots are imaginary, the response is underdamped or oscillatory (Fig. 4.10). If the roots are equal, the response is critically damped.

the *undamped natural frequency* as

$$\omega_n = \sqrt{1/JK} \tag{26-10}$$

Another useful quantity is the ratio of the actual damping parameter D to D_c, the value for critical damping. Since for critical damping the discriminant is zero, the corresponding value of D_c is $2\sqrt{J/K}$. The *damping ratio* ζ (zeta) is defined as

$$\zeta = \frac{D}{D_c} = \frac{D}{2\sqrt{J/K}} \tag{26-11}$$

The *natural frequency* ω of an underdamped system can be expressed in terms of ω_n and ζ. In this oscillatory case (see Eq. 4-18c) the discriminant is negative and the natural frequency is defined by

$$\omega = \sqrt{\frac{1}{JK} - \frac{D^2}{4J^2}} = \frac{1}{\sqrt{JK}}\sqrt{1 - \frac{D^2}{4J/K}} = \omega_n \sqrt{1 - \zeta^2} \tag{26-12}$$

In terms of these definitions, the transfer function of a general second-order system (see Example 3) becomes

$$T(s) = \frac{Y(s)}{X(s)} = \frac{k}{s^2 + 2\omega_n\zeta s + \omega_n^2} \tag{26-13}$$

Let us now determine the response $y(t)$ to a step of magnitude X for a system characterized by the transfer function of Eq. 26-13. By inspection, the forced response is

$$y_f = \frac{k}{\omega_n^2}X = Y \tag{26-14}$$

The roots of the denominator are determined from

$$s^2 + 2\omega_n\zeta s + \omega_n^2 = 0 \tag{26-15}$$

whence

$$s_1, s_2 = -\omega_n(\zeta \pm \sqrt{\zeta^2 - 1}) \tag{26-16}$$

The complete response is of the form

$$y(t) = \frac{k}{\omega_n^2}X + A_1e^{s_1t} + A_2e^{s_2t} \tag{26-17}$$

The coefficients A_1 and A_2 can be determined from the initial conditions assuming that at $t = 0^+$, $y = 0$ and $dy/dt = 0$. After several algebraic steps, the normalized response is found to be†

$$\frac{y}{Y} = 1 - \frac{\zeta + \sqrt{\zeta^2 - 1}}{2\sqrt{\zeta^2 - 1}}e^{-(\zeta - \sqrt{\zeta^2-1})\omega_n t} + \frac{\zeta - \sqrt{\zeta^2 - 1}}{2\sqrt{\zeta^2 - 1}}e^{-(\zeta + \sqrt{\zeta^2-1})\omega_n t} \tag{26-18}$$

Choosing y/Y and $\omega_n t$ as the normalized variables, the response is as plotted in Fig. 26.9.

These curves, developed by Gordon Brown, are useful in predicting the response of a system whose parameters are known or in estimating the parameters of a system whose response is known. If the damping ratio is small, say, $\zeta = 0.2$, the output variable rises rapidly to the steady-state value, *overshoots* by 50%, and oscillates several times before the natural response component becomes negligible. If the damping ratio is large, say, $\zeta = 1.5$, the output variable approaches the steady-state value very slowly. If $\zeta = 1.0$, the damping is at the critical value and the output variable approaches the steady-state value at the

---

† For $\zeta = 1$ this form of the equation is indeterminate; the response with critical damping is $y/Y = 1 - (1 + \omega_n t)e^{-\omega_n t}$. For $\zeta < 1$, it is more convenient to express the response as a damped sinusoid.

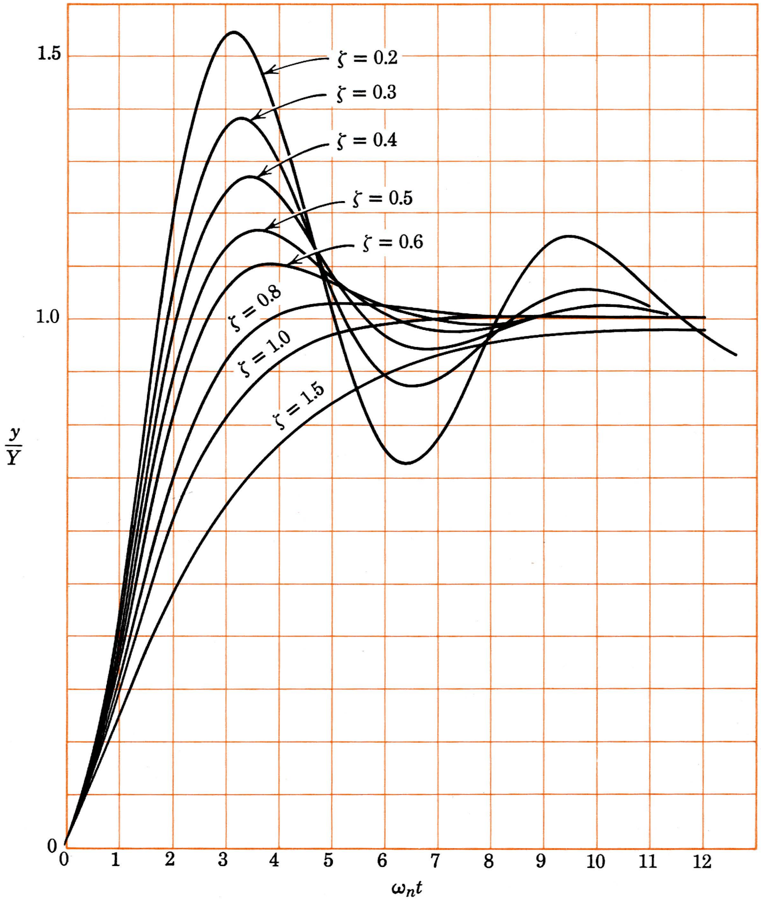


Figure 26.9 Normalized response curves for second-order systems.

maximum rate possible without overshoot. If rapid response is desired and a small amount of overshoot is tolerable, a design value of $\zeta = 0.8$ might be specified.

System Time Constants

One of the virtues of the transfer function is the variety of useful ways in which it can be interpreted. In several examples we have seen that the form of the step response of a system is determined by the poles of $T(s)$. The *pole*

Example 4

A d'Arsonval movement to be used in a pen recorder has a rotational inertia $J = 2.5 \times 10^{-8}$ kg-m^2. It is to have a damping ratio $\zeta = 0.8$ and an undamped natural frequency $\omega_n = 200$ rad/s. Specify the spring compliance, determine the developed torque required for a full-scale deflection of 1.2 radian, and estimate the time required for the deflection to reach 0.98 of the steady-state value.

Assuming that the movement is represented by the circuit model of Fig. 26.8a (Example 3), this second-order system is characterized by the transfer function of Eq. 26-13. From Eq. 26-10 the necessary spring compliance is

$$K = \frac{1}{\omega_n^2 J} = \frac{1}{4 \times 10^4 \times 2.5 \times 10^{-8}} = 10^3 \text{ rad/N-m}$$

For a full-scale deflection of 1.2 rad, the torque required is

$$\tau_d = \frac{\Theta}{K} = \frac{1.2 \text{ rad}}{10^3 \text{ rad/N-m}} = 1.2 \times 10^{-3} \text{ N-m}$$

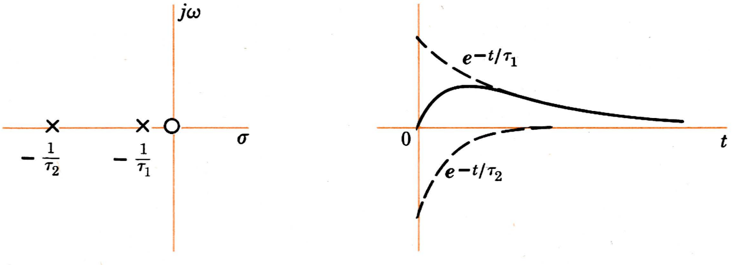
For $\zeta = 0.8$, Fig. 26.9 indicates that $y/Y = 0.98$ when $\omega_n t \cong 3.6$ rad. Therefore,

$$t \cong \frac{3.6}{\omega_n} = \frac{3.6}{200} = 0.018 \text{ s}$$

and *zero* locations are explicit when the transfer function is written in the form

$$T(s) = \frac{k(s + \beta)}{(s + \alpha_1)(s + \alpha_2)} \tag{26-19}$$

For the special case of real negative poles and a zero at the origin ($\beta = 0$), the pole-zero diagram is as shown in Fig. 26.10.



(a) Pole–zero diagram (b) System time constants

Figure 26.10 Interpretations of the transfer function $T(s) = k's/(s\tau_1 + 1)(s\tau_2 + 1)$.

Rewriting Eq. 26-19 (for $\beta = 0$) as

$$T(s) = \frac{ks/\alpha_1\alpha_2}{(s/\alpha_1 + 1)(s/\alpha_2 + 1)} = \frac{k's}{(s\tau_1 + 1)(s\tau_2 + 1)} \tag{26-20}$$

focuses attention on a different set of system characteristics. Here τ_1 and τ_2 are the *time constants* of the natural response terms. The relation of the time constants to the step response is shown in Fig. 26.10b.

Example 5

Determine the transfer function of a first-order servomechanism to meet the following specifications:

Maximum displacement:	25°
Error: no greater than 0.5° after 2 s	
Static gain:	10

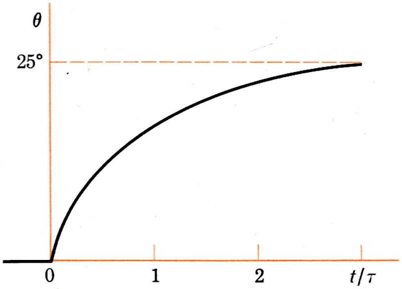


Figure 26.11 Deriving the transfer function of a servomechanism.

The desired behavior is shown in Fig. 26.11 and we conclude that the form of the necessary transfer function is (Eq. 26-3)

$$T(s) = \frac{Y(s)}{X(s)} = \frac{k}{s+\alpha} = \frac{k\tau}{s\tau + 1}$$

to provide a step response

$$\frac{y}{Y} = 1 - e^{-t/\tau}$$

An error of 0.5° corresponds to 2% of the maximum displacement, therefore,

$$1 - 0.98 = 0.02 = e^{-t/\tau}$$

and the time constant should be

$$\tau = \frac{-t}{\ln 0.02} = \frac{-2}{-3.9} \cong 0.51 \text{ seconds}$$

For a gain of $Y/X = 10$ at $s = 0$, $k\tau = 10$; therefore the desired system transfer function is

$$T(s) = \frac{10}{0.51s + 1}$$

System Frequency Response

Another useful interpretation of the transfer function is in terms of the steady-state response, i.e., amplitude and phase angle, as a function of frequency. In some cases, a laboratory measurement of frequency response is the best method of determining the system characteristics. Given a transfer function, the frequency response is obtained directly by substituting $s = j\omega$. For the system described by $T(s) = 1/(s\tau + 1)$, the frequency response is

$$T(j\omega) = \mathbf{T} = \frac{1}{1 + j\omega\tau} = \frac{1}{\sqrt{1 + (\omega\tau)^2}} \underline{/\tan^{-1} - \omega\tau} \qquad (26\text{-}21\text{a})$$

with a breakpoint at $\omega\tau = 1$. For the system described by $T(s) = s\tau/(s\tau + 1)$, the frequency response is

$$T(j\omega) = \mathbf{T} = \frac{j\omega\tau}{1 + j\omega\tau} = \frac{1}{1 - j\dfrac{1}{\omega\tau}} = \frac{1}{\sqrt{1 + \left(\dfrac{\omega_1}{\omega}\right)^2}} \underline{/\tan^{-1}(\omega_1/\omega)} \quad (26\text{-}21b)$$

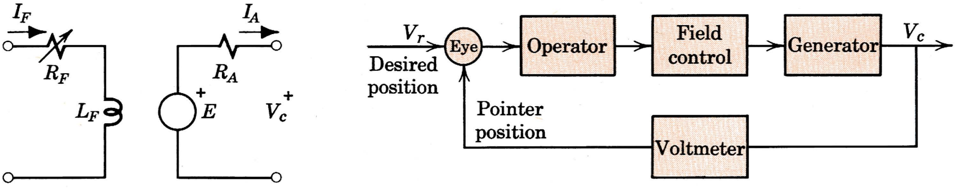
where the breakpoint is defined by $\omega_1\tau_1 = 1$ or $\omega_1 = 1/\tau_1$ in the nomenclature used with amplifiers. If the transfer function of the d'Arsonval movement of Fig. 26.6 has complex roots, the frequency response is a little more complicated but the procedure is the same. (See Problem 3.)

Feedback Control Systems

With an understanding of the relationship between transfer functions and dynamic responses, we are ready to consider control systems with feedback. Our aim is to learn how they operate, how they can be used to improve performance, and how system stability can be ensured. Because of our experience with circuits and electromechanical devices we choose an electrical system as an illustration.

A Voltage Regulator

The purpose of a voltage regulating system is to hold the controlled voltage within specified limits in spite of changes in load or other operating conditions. The output voltage of the unregulated, separately excited, constant-speed, dc generator in Fig. 26.12a varies with changes in load current because of the armature-resistance voltage drop. The variation can be reduced by "closing the loop" through an operator with his eye on the voltmeter and hand on the field rheostat (Fig. 26.12b). The "controller" notes the "error" between the voltmeter pointer position and the "reference" and takes the necessary "action" to reduce the error to zero.



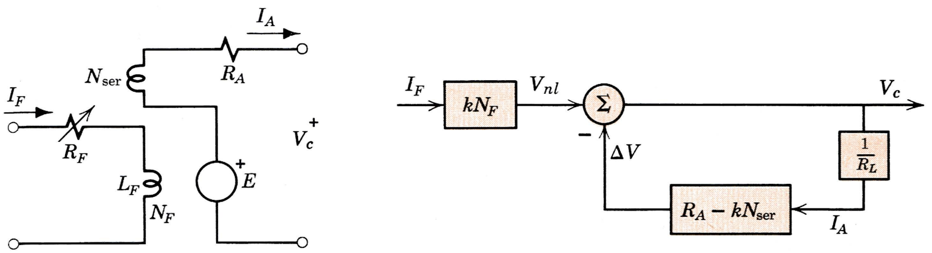
(a) Circuit model (b) Block diagram

Figure 26.12 Operation of a manual voltage regulator.

In the constant-speed compound generator of Fig. 26.13 the effect of a load change is fed back by means of the series winding (see Example 3 in Chapter 24). If operation is on the linear portion of the magnetization curve and changes are slow, the governing equation is

$$\begin{aligned} V_c &= E - I_A R_A = k(N_F I_F + N_{\text{ser}} I_A) - I_A R_A \\ &= V_{nl} - \Delta V = kN_F I_F - I_A(R_A - kN_{\text{ser}}) \end{aligned} \quad (26\text{-}22)$$

where N is the number of turns.

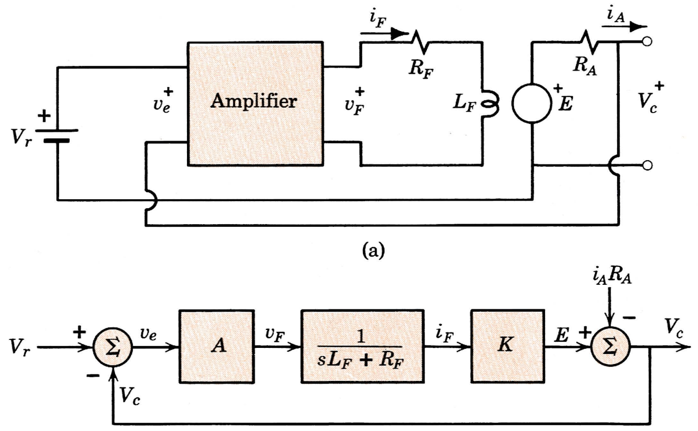

I_A
N_{ser}
R_A
I_F
R_F
V_c
L_F
N_F
E
I_F
kN_F
V_{nl}
Σ
ΔV
V_c
$\frac{1}{R_L}$
$R_A - kN_{ser}$
I_A

(a) Circuit model (b) Block diagram

Figure 26.13 Steady-state operation of a compound generator.

The corresponding block diagram is shown in Fig. 26.13b. For flat compounding, the term $R_A = kN_{ser}$ is made equal to zero, $\Delta V = 0$, and $V_c = V_{nl}$. But there are two inadequacies in this system. First, the feedback loop contains an element R_A that changes with temperature (and therefore with I_A) and factor k that changes with magnetic flux (and therefore with I_A). Second, there is no amplification in the loop and the total error is incorporated in the output.

The voltage regulator in Fig. 26.14 represents an improvement. Here the output voltage is compared to a reference, and the difference is amplified and used to decrease the discrepancy. In anticipation of a rapid response, we show the variables as instantaneous values. The operation is: A sudden change in load current reduces output voltage V_c and greatly increases the error voltage $v_e = V_r - V_c$; the error voltage is then amplified and applied to the field circuit, tending to increase i_F and restore the output voltage. For linear operation, the


i_F
i_A
R_F
R_A
v_e
Amplifier
v_F
L_F
E
V_c
V_r
(a)
i_AR_A
V_r
Σ
v_e
A
v_F
$\frac{1}{sL_F + R_F}$
i_F
K
E
Σ
V_c
V_c

(b)

Figure 26.14 Wiring and block diagrams for a voltage regulator.

governing equations are:

$$v_e = V_r - V_c \qquad E = Ki_F$$
$$v_F = Av_e \qquad V_c = E - i_A R_A \tag{26-23}$$
$$v_F = L_F \frac{di_F}{dt} + R_F i_F$$

When each of these relations is represented by an element, the block diagram is as shown.

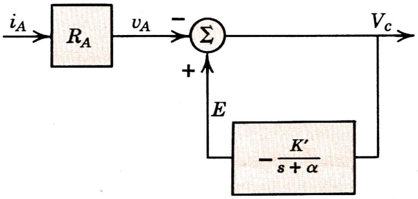


Figure 26.15 A voltage regulator with a disturbance input.

In this representation the voltage drop $v_A = i_A R_A$ is introduced as a disturbance to the system. One approach to evaluating the performance of the regulator is to determine its response to a step in load current I_A. In this situation the reference voltage V_r is constant and the important relation is between I_A and V_c. To focus attention on this relation, the block diagram is redrawn as shown in Fig. 26.15. Here the forward transfer function $G(s) = +1$. With V_r constant, $v_e = -\Delta V_c$ and the feedback transfer function relating ΔE to ΔV_c is

$$H(s) = -\frac{KA}{sL_F + R_F} = -\frac{KA/L_F}{s + R_F/L_F} = -\frac{K'}{s+\alpha} \tag{26-24}$$

where K' is a new constant and $\alpha = R_F/L_F$. For this configuration, the overall gain with feedback (Eq. 26-1) is

$$\frac{V_c(s)}{I_A(s)} = -R_A G_F(s) = R_A \frac{-1}{1 - H(s)} = \frac{-R_A}{1 + [K'/(s+\alpha)]} = \frac{-R_A(s+\alpha)}{s + \alpha + K'} \tag{26-25}$$

With the transfer function known, the step response of the system is determined from forced and natural components and initial conditions. The forced component of the change in controlled voltage is (for $s = 0$)

$$v_f = -\frac{R_A \alpha}{\alpha + K'} I_A = -\frac{R_A R_F/L_F}{R_F/L_F + KA/L_F} I_A = -\frac{R_A R_F I_A}{R_F + KA} \tag{26-26}$$

The natural response is defined by the pole of $G_F(s)$ and

$$v_n = A_n e^{-(\alpha + K')t} \tag{26-27}$$

The performance of the regulator is revealed by the variation of the total controlled voltage V_c as a function of time after the introduction of the disturbance I_A. In general,

$$V_c = V_{nl} + \Delta V_c = V_{nl} + v_f + v_n$$
$$= V_{nl} - \frac{R_A R_F I_A}{R_F + KA} + A_n e^{-(\alpha + K')t} \tag{26-28}$$

If speed is constant as assumed, the generated voltage E cannot suddenly change because i_F cannot suddenly change. (Why not?) Therefore, if the no-load output voltage is V_{nl} and a current I_A is suddenly drawn at $t = 0$, at $t = 0^+$

$$V_c = V_{nl} - I_A R_A = V_{nl} - \frac{R_A R_F I_A}{R_F + KA} + A_n e^0 \tag{26-29}$$

Solving Eq. 26-29,

$$A_n = -I_A R_A \left(1 - \frac{R_F}{R_F + KA}\right) \tag{26-30}$$

and therefore,

$$V_c = V_{nl} - \frac{R_A R_F I_A}{R_F + KA} - I_A R_A \left(1 - \frac{R_F}{R_F + KA}\right) e^{-(R_F + KA)t/L_F} \tag{26-31}$$

The effect of the various parameters is illustrated in the following example.

Example 6

The shunt generator in Fig. 26.14 ($R_A = 1\ \Omega$, $R_F = 100\ \Omega$, $L_F = 20$ H) is rated at 100 V and 10 A. The magnetization curve can be approximated by the relation $E = 100 I_F$ V where I_F is in A.

Predict the effect of a load current of 10 A on the output voltage of the generator operating without feedback.

Predict the effect of the same load current with the feedback circuit shown incorporating an amplifier with voltage gain $A = 100$.

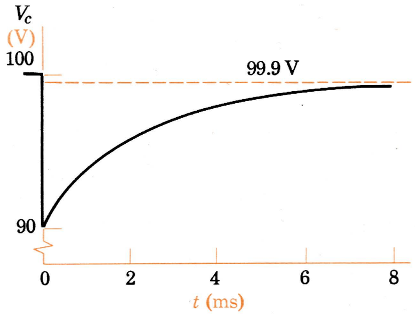


Figure 26.16 Step response of a voltage regulator.

With the field current adjusted for rated voltage at no load but without feedback, the output voltage under load is

$$V = E - I_A R_A = 100 - 10 \times 1 = 90 \text{ V}$$

The application of this load will result in a 10% voltage drop.

With feedback, Eq. 26-31 applies. Here

$$\frac{R_F}{R_F + KA} = \frac{100}{100 + 100 \times 100} \cong 0.01$$

and

$$\frac{R_F + KA}{L_F} = \frac{100 + 100 \times 100}{20} \cong 500$$

hence

$$V_c = V_{nl} - 0.01 I_A R_A - I_A R_A (1 - 0.01) e^{-500t}$$

For $v_c = V_{nl} = 100$ V and $I_A R_A = 10 \times 1 = 10$ V introduced at $t = 0$,

$$V_c = 100 - 0.01 \times 10 - 10(0.99) e^{-500t}$$

or

$$V_c = 99.9 - 9.9 e^{-500t}$$

As shown in Fig. 26.16, the voltage will drop abruptly to 90 V, but within a few milliseconds it has been increased to 99.9 V. The application of a 10-A load results in a 0.1% voltage drop compared to the 10% voltage drop in the same generator without feedback.

Improving Response

An open-loop system is "ignorant of its own output." In contrast, a closed-loop system is continually reminded of what is occurring at the output terminals. The principal virtue of feedback in amplifiers (Chapter 18) is the ability to provide satisfactory operation under widely ranging conditions. In control applications, feedback can greatly improve system performance by reducing errors or by extending the frequency response or by reducing the time constant of a system component.

In the voltage regulator of Example 6, the difference between the controlled voltage and the reference voltage, the *error,* is the input to the amplifier. The purpose of the system is to drive this error to zero. On the other hand, some error must exist to provide the necessary input. As an illustration, consider an aircraft rudder control. The difference between the actual rudder position and that called for by the pilot is sensed and amplified. The output of the amplifier is the input to a motor that provides the torque to actuate the rudder. In a turn, considerable torque is required and therefore there must be some steady-state error. This error is kept small by using high amplification.

It is characteristic of control systems that the input consists of random variations. These may be slow or fast corresponding to low and high frequencies. The ability of the system to respond satisfactorily to random variations can be evaluated in terms of its steady-state frequency response as shown by a Bode plot (Fig. 9.10c). The frequency response of a mechanical system (such as the d'Arsonval movement in Example 4) can be extended by using feedback. The technique is similar to that employed to extend the frequency response of an amplifier (see Example 8 in Chapter 18). In an amplifier, increased bandwidth is obtained at the expense of gain. In a servomechanism, extended frequency response is obtained at the sacrifice of simplicity. Some form of amplification is always required.

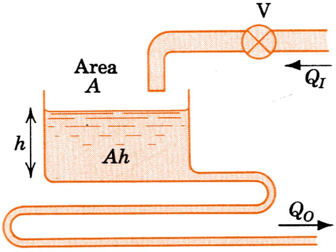
The speed of response is another limiting factor in the performance of control systems. It is desirable that the system have a low dynamic error as well as a low steady-state error. How can the time constant of a physical element be reduced using feedback? In Example 6, the response of the basic generator to a change in field voltage v_F was defined by the time constant $\tau = L_F/R_F = 20/100 = 0.2$ s. With feedback, the time constant is reduced to $\tau_F = L_F/(R_F + KA) = 20/(100 + 100 \times 100) \cong 0.002$ s; as expected, the time constant is reduced by a factor that can be interpreted as $(1 + GH)$. As an illustration of the reduction of time constants in another field consider Example 7.

Stability

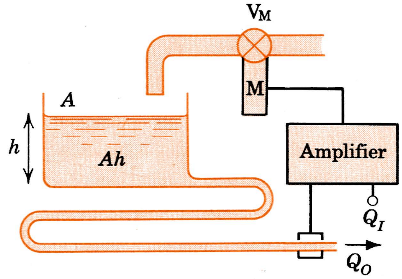
We see that feedback can improve the performance of a system by making it faster, more accurate, or responsive over a wider range of frequencies. In each case amplification is necessary, and the degree of improvement is related to the

Example 7

In a chemical plant, water flows into an open mixing tank (Fig. 26.17a) and then out the pipe at the bottom. The rate of flow is governed by an ON–OFF inlet valve V; the head h in the tank increases until the rate of flow out Q_O (directly proportional to h in this case of laminar flow) is equal to the flow rate in Q_I. In normal operation, the tank is frequently drained and refilled. Devise a system to reduce the delay in reaching steady-state conditions after draining.



(a)



(b)

With the aid of feedback and some auxiliary equipment, the time constant of this unit can be reduced. For the original arrangement, the governing equation is

$$Q_I = Q_O + \frac{d(Ah)}{dt} = kh + A\frac{dh}{dt}$$

where Ah is the volume of water in the tank.

The transfer function is

$$T(s) = \frac{H(s)}{Q_I(s)} = \frac{1}{sA + k}$$

with a time constant $\tau = A/k$.

In Fig. 26.17b, Q_O is sensed and converted to an electrical signal that is compared to a reference signal corresponding to the original rate Q_I. The error $Q_O - Q_I$ is amplified and the amplified signal controls the motorized valve V_M. As shown in the block diagram (Fig. 26.17c), the overall gain of the amplifier and valve is G. The system response with feedback (Eq. 26-1) is

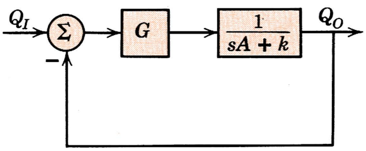
$$Q_O = kh = Q_I\frac{G/(sA + k)}{1 + G/(sA + k)} = Q_I\frac{G}{sA + k + G}$$

The new transfer function is

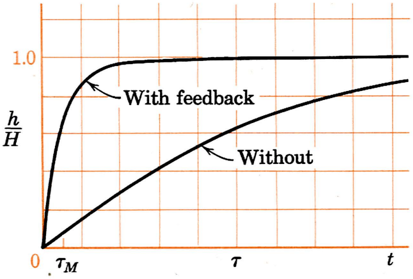
$$T_M(s) = \frac{H(s)}{Q_I(s)} = \frac{G/k}{sA + (k + G)}$$

with a time constant $\tau_M = A/(k + G)$.

If, by proper design of the amplifier and valve, a gain $G = 9k$ is obtained, the new time constant $\tau_M = \tau/10$. The corresponding variations of water level with time are shown in Fig. 26.17d.



(c)



(d)

Figure 26.17 Time constant reduction by using feedback.

amount of amplification. The combination of amplification and feedback may also make the system unstable and therefore ineffective.

The mathematical basis for instability is revealed in the closed-loop gain equation where

$$\mathbf{G}_F = \frac{\mathbf{G}}{1 + \mathbf{GH}} \tag{26-32}$$

If for any frequency the denominator is zero, the gain increases without limit and the system is unstable. The denominator is zero if the open-loop transfer function **GH** is just equal to $-1 + j0$. A possible approach is to plot the amplitude and phase of **GH** as functions of frequency as in Fig. 9.10c. If the gain is 1 at a phase shift of ±180°, the system is unstable.†

An alternative statement is that the system is unstable if any poles of the overall transfer function

$$T(s) = \frac{G(s)}{1 + G(s)H(s)} \tag{26-33}$$

are on the $j\omega$ axis or in the right half of the complex frequency plane. (Where are the poles for the functions whose responses are given in Fig. 26.4?)

In designing automatic feedback control systems, much of the engineer's work is concerned with stability analysis and there are many books devoted to the subject. After the system has been designed to perform the desired function, precautions must be taken to ensure stability. Amplifier gains, time constants, gear ratios, and damping ratios must be selected so that the poles of the transfer function are kept out of the right half of the s plane. Sometimes compensating networks are placed in series with system components to cancel out undesirable poles. Auxiliary feedback loops may also be used to provide stability.

The basic question is whether the denominator of $T(s)$ has any roots in the right half of the s plane. If the denominator is known in factored form (see Eq. 4-34) this question is readily answered. In general, however, $G(s)$ and $H(s)$ are polynomials in s and finding the roots is a tedious operation. From the design engineer's viewpoint, the practical problem is to determine the effect on stability of variations in system parameters. To optimize system design, he must provide adequate stability with the specified performance at the least cost (in complexity, reliability, weight, or dollars). Various methods have been developed to provide convenient solutions to this problem in the practical case where the system transfer functions are quite complicated.

One of the most widely used techniques in system analysis is the *root-locus method* developed by W. R. Evans in 1948. By this technique, the poles of the transfer function are located graphically as the system parameters are varied. The value of such a graphical method can be illustrated by a specific numerical example. For the system defined by

---

† For a discussion of the Nyquist stability criterion see any of the references at the end of this chapter.

$$G(s) = \frac{K(s+2)}{s(s+1)(s+4)(s+10)} \qquad H(s) = 1 \tag{26-34}$$

the overall transfer function is

$$T(s) = \frac{G(s)}{1+G(s)} = \frac{K(s+2)}{K(s+2)+s(s+1)(s+4)(s+10)} \tag{26-35}$$

The transfer function has a zero at $s = -2$, and for $K = 0$, poles at $s = 0, -1, -4$, and -10. Using the root-locus technique,† the loci of all points satisfying the relation

$$\underline{/G(s)} = \pm 180° + n(360°) \tag{26-36}$$

are as shown in Fig. 26.18. The roots move along the loci as K is varied. As K increases from zero, the poles which start at $s = 0$ and $s = -1$ come together and then diverge along the curved locus. For values of K that result in poles close to the $j\omega$ axis, the step response of the system is highly oscillatory. The

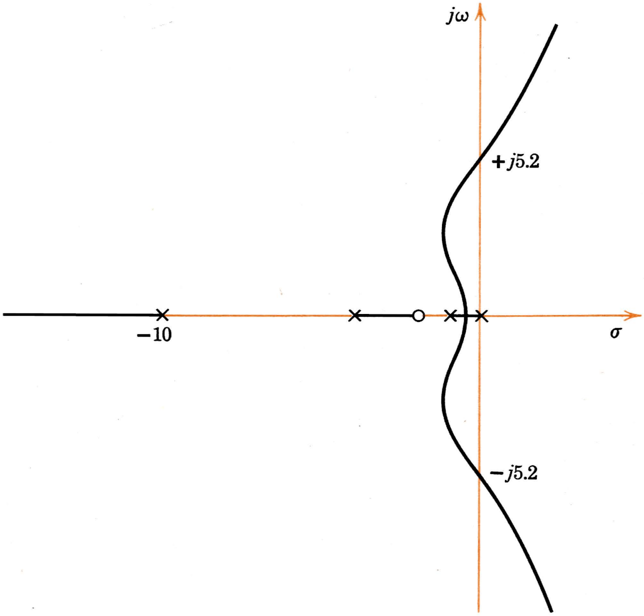


Figure 26.18 All points satisfying $\underline{/G(s)} = \pm 180°, \pm 540°$. (From R. N. Clark, *Introduction Automatic Control Systems*, John Wiley & Sons, New York, 1962, by permission.)

---

† See p. 192 of the Clark reference at the end of this chapter.

critical value of K for stability occurs where the poles are on the $j\omega$ axis. On the basis of the root-locus plot, an experienced systems engineer can decide on an appropriate value of K.

SUMMARY

- Automatic feedback control systems employ transducers, error detectors, amplifiers, actuators, and feedback elements to provide rapid and accurate control.

  Control system analysis is based on circuit principles, device characteristics, and system techniques.

- In comparison to an open-loop system, a closed-loop system is capable of greater accuracy over a wider range of conditions using less precise control elements.

  The control engineer's problem is to provide the necessary performance with adequate stability at minimum cost.

  System performance can be predicted on the basis of step response.

- The transfer function is the ratio of two exponential functions of time; it is derived and interpreted just as an immittance function.

  The dynamic response of a system element is obtained from the transfer function.

  System behavior is defined by the overall transfer function.

- The order of a system is defined by the highest power of s in the denominator of the transfer function. The step response is:

  $$y(t) = A_0 + Ae^{st} \text{ for a first-order system, and}$$
  $$y(t) = A_0 + A_1e^{s_1t} + A_2e^{s_2t} \text{ for a second-order system.}$$

  Normalized curves display the effect on response of damping ratio ζ.

- Feedback can improve control system performance by reducing errors, by extending the frequency response, or by reducing time constants.

- The combination of amplification and feedback may result in instability. Much of the systems engineer's work is concerned with stability analysis.

REFERENCES

1. J. G. Truxal, *Introductory System Engineering*, McGraw-Hill Book Co., New York, 1972.
2. R. N. Clark, *Introduction to Automatic Control Systems*, John Wiley & Sons, New York, 1962.
3. J. J. D'Azzo and C. H. Houpis, *Feedback Control System Analysis and Synthesis*, McGraw-Hill Book Co., New York, 1960.

REVIEW QUESTIONS

1. Why don't power inputs appear in a control system block diagram?
2. Explain in system terminology how highway traffic speed is regulated.
3. Explain in system terminology how body temperature is regulated.
4. What information on system performance is provided by step response?
5. Define "transform," "transfer function," and "step response."
6. Outline the procedure for determining step response from the transfer function.
7. What information is obtained from the poles of $T(s)$? The zeros?
8. What form of step response is indicated by complex poles of $T(s)$?
9. Define ζ. What are its dimensions? Why is it useful?
10. What value of damping ratio results in a 20% overshoot?

EXERCISES

1. In an automobile "Cruise Control" system, the desired speed is manually set into the comparator. The electrical output of the comparator passes through a transistorized amplifier to the solenoid valve of a power unit. The power unit moves the engine throttle which drives the driveshaft. The speedometer is connected to the driveshaft and its reading is introduced into the comparator.
   - (a) Draw a block diagram of this automatic control system.
   - (b) Identify the controlled output, the feedback element, the error detector, and the reference input.
2. In an antenna-positioning system, the input direction dial and the antenna are connected to similar rotating potentiometers across a common voltage V. The voltage difference is fed into a corrective network, then into a power amplifier that drives a motor that turns the antenna through a set of gears.
   - (a) Draw a block diagram of this automatic control system.
   - (b) Identify the controlled output, the feedback element, the error detector, and the reference input.
3. For the flywheel of Fig. 26.19,
   - (a) Define and derive the transfer function.
   - (b) Sketch the pole-zero diagram of $T(s)$.
   - (c) Sketch the step response.
4. Given $T(s) = (s + a)/(s + 2a)$,
   - (a) Sketch the pole-zero diagram.
   - (b) Determine the step response.
5. The reaction time of an automobile driver, $T_R = 0.1$ s, represents a pure time delay in that $x_2(t) = x_1(t - 0.1)$.

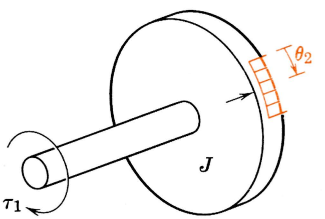


Figure 26.19

   - (a) Derive the transfer function $X_2(s)/X_1(s)$,
   - (b) Derive the transfer function $T(j\omega)$.
6. Determine the response to a step of amplitude 10 (no initial energy storage) given:
   - (a) $T(s) = 2s/(s + 5)$
   - (b) $T(s) = 5/(s + 2)$
7. Determine the transfer function given the unit step response:
   - (a) $x(t) = 5 - 5e^{-2t}$
   - (b) $i(t) = 6e^{-t/50}$
8. Derive the transfer function $T(j\omega)$ relating displacement x to force f for the system of Fig. 9.16. For $M = 0.1$ kg and $K = 0.4$ m/N, sketch the amplitude A and the phase angle ϕ as functions of frequency ω.
9. An amplidyne is a rotating machine in which the voltage generated in one winding provides field current for a second winding. Such a power amplifier can be represented by the circuit in Fig. 26.20.

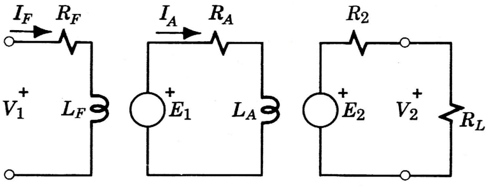


Figure 26.20

(a) Assuming linear magnetic fields, draw a block diagram of the amplidyne.
(b) Derive the transfer function.
(c) Determine the forced response to a step input ΔV.
(d) Determine the *form* of the natural step response.

10. (a) Explain why the natural response of a system is defined by the poles of the transfer function rather than the zeros.
(b) Given $T(s) = (s + a)/(s^2 + sb)$, write an expression for the natural response.

11. A shunt motor ($J = 2 \times 10^{-6}$ kg-m^2 and $D = 8 \times 10^{-5}$ N-m-s/rad) is operated with constant armature current. The inductance of the field circuit is 200 mH and the resistance is 20 Ω. Derive the transfer function relating shaft velocity to field voltage and predict (and sketch) the step response.

12. In Fig. 26.21, a shunt motor with constant field excitation drives a spring-loaded shaft ($K = 3000$ rad/N-m). The equivalent friction torque of the motor is $D' = 3 \times 10^{-5}$ N-m-s/rad and the motor inertia is $J = 6 \times 10^{-6}$ kg-m^2. Derive the transfer function relating angle θ and input voltage v and sketch the expected step response.

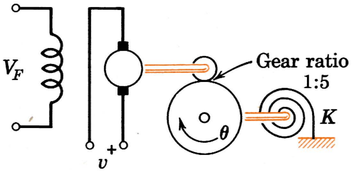


Figure 26.21

13. Repeat Exercise 12 where K is reduced to 300 rad/N-m.

14. The frequency response of a transfer function $T(j\omega)$ is given in Fig. 26.22.
(a) Determine the transfer function.
(b) Determine the step response, assuming no initial energy storage.

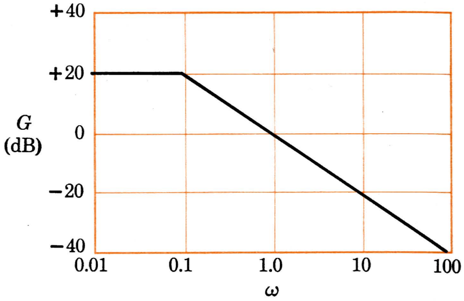


Figure 26.22

15. In the voltage regulator of Fig. 26.23, the output of the constant-speed generator is $E = 100 i_F$ V where i_F is in amperes. The dc amplifier gain is $K = -100$, $R_F = 50\ \Omega$, $L_F = 10$ H, and $R_A = 1\ \Omega$.
(a) Predict V for $R_L = 1\ \Omega$.
(b) Predict and sketch $V(t)$ if R_L is suddenly changed to 5 Ω.

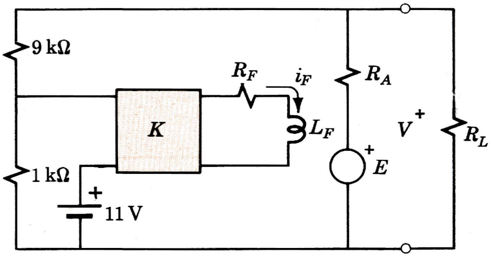


Figure 26.23

PROBLEMS

1. An ac servomotor has a speed-torque curve approximated by a straight line with intercepts $\Omega/\Omega_s = 1.0$ and $T/T_R = 2.0$. (See Fig. 25.22.) A handbook states: "The time t_a required for a servomotor to accelerate from rest to 63% of its no-load speed n_o (rpm) is $t_a = (J\,n_o/T_s) \times 1.483 \times 10^{-6}$ S, where J = rotor inertia (g − cm^2) and T_s = starting torque (oz − in.)." Stating any necessary assumptions, derive this relationship.

2. The d'Arsonval meter of Fig. 26.6 has a steady deflection θ_o for a dc voltage V. The damping ratio ζ is 0.5.
(a) Describe the response to a suddenly applied dc voltage V.
(b) Describe the frequency response, i.e., deflection amplitude as a function of the frequency of the applied voltage.
(c) Determine and plot the normalized frequency response θ/θ_o vs ω/ω_o where $\omega_o = \sqrt{1/JK}$.

Chapter 27

Instruments and Instrumentation Systems

Electrical Indicating Instruments
Electronic Indicating Instruments
Instrumentation Transducers
Instrumentation Systems

All branches of experimental engineering and science depend on instrumentation. Instruments and instrumentation systems are used for detecting, observing, measuring, controlling, computing, communicating, and displaying physical quantities. In a broad sense, instruments extend human sensing abilities by measuring more accurately, or more rapidly, or over greater ranges, or at greater distances, or by measuring quantities to which human beings are insensitive.

From the engineer's viewpoint, a knowledge of instrumentation principles and techniques is essential for studying the characteristics of existing devices in a test facility, for designing automatic operating and processing systems in an industrial plant, and for discovering new truths in a research laboratory. In any instrumentation system there is a flow of information in the form of signals. Because of the ease and precision with which electrical signals are processed, it is customary to convert physical variables into electrical form for measuring, communicating, and computing. The emphasis here, therefore, is on electrical instrumentation.

The information on circuits, electronic and electromechanical devices, amplifiers, and feedback systems presented in this book provides an adequate background for the study of almost any instrumentation system. Our purpose here is to use this background in providing an introduction to instrumentation principles and techniques that will be valuable in undergraduate laboratory work, in graduate research projects, and as preparation for professional employment. The sophisticated instrumentation used in the various branches of science and engineering is highly specialized; here we can only look at a few of the more important instruments and methods.

First, we study the basic electrical indicating instruments and their characteristics. Next, we consider the more sophisticated electronic instruments that provide greater accuracy, flexibility, or convenience. Then we examine some common transducers for converting physical variables into electrical signals for processing. Finally, we look at some of the characteristics of comprehensive instrumentation systems.

Electrical Indicating Instruments

The d'Arsonval moving-coil mechanism (see Fig. 3.22a) is a versatile device for displaying electrical currents or other physical variables that are directly related to currents. In *indicating* instruments the moving coil carries a needle that gives a temporary indication of the current value; in *recording* instruments the moving coil actuates a pen or stylus so as to leave a permanent record. The *taut-band suspension,* in which the movement is supported by two metal bands that carry the current and provide restoring torque, eliminates bearing friction and provides highly reproducible measurements.

DC Measurements

An ordinary portable instrument may have a movement requiring about 50 μW for full-scale deflection. If many turns of fine wire are used on the moving coil, the resistance of the movement is high, but the current required for full-scale deflection, i.e., the current *sensitivity,* is low. Conversely, the coil may be of fewer turns and lower resistance, but a larger current is required for full-scale deflection. As pointed out in Chapter 23 (see Fig. 23.10), the same movement can be used for measuring current and voltage.

The Ammeter. Current in a branch is measured by an ammeter connected in series with the branch. Because the ammeter has some internal resistance, insertion of the ammeter may change the current in the branch from its original value. A useful model of a real ammeter consists of an ideal resistanceless ammeter in series with a resistance equal to the meter resistance (see Example 1). Knowing the characteristics of the circuit, we can predict the effect of the ammeter or we can specify the maximum allowable meter resistance for a negligible effect.

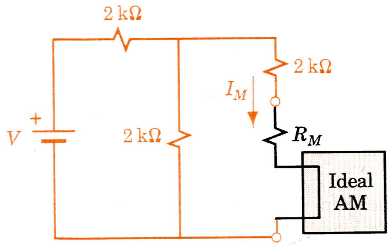
In Example 1, the insertion of the instrument introduced an error of greater than 6%. To keep this error less than 1%, Eq. 27-2 indicates that $R_M/R' \leq 0.01$ or $R_M \leq 0.01R' = 32\,\Omega$. A typical value for a portable 1-mA instrument might be around 25 Ω.

A circuit for a multirange ammeter is shown in Fig. 27.2a. Four shunts are available to give four ammeter ranges. A make-before-break selector switch selects the appropriate range. Initial current readings should always be made on the highest current range; a lower range can then be selected to obtain an easily read deflection.

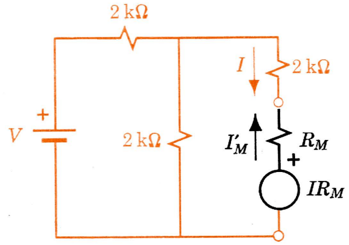
The Voltmeter. Voltage across a branch is measured by a voltmeter in parallel with the branch. The d'Arsonval movement is converted to a voltmeter by

Example 1

The current in a branch of the circuit in Fig. 27.1a is to be measured by inserting a milliammeter with an estimated internal resistance $R_M = 200\ \Omega$. The meter reads 0.937 mA. Predict the current without the ammeter.



(a) Connection



(b) Compensation

Figure 27.1 Effect of ammeter on current.

The compensation theorem states†:

The effect on any network of increasing the resistance of a branch (originally carrying a current I) by an amount R_M is the same as the effect of inserting an emf IR_M in the modified branch.

As shown in Fig. 27.1b, the voltage drop due to the insertion of the meter resistance R_M is just compensated by the addition of a fictitious emf IR_M. The increment of current due to the compensating emf is, by the theorem of superposition,

$$I'_M = \frac{IR_M}{R'} \tag{27-1}$$

where R' is the resistance seen by the ideal ammeter. The actual ammeter reading is

$$I_M = I - I'_M = I\left(1 - \frac{R_M}{R'}\right) = I\left(\frac{R' - R_M}{R'}\right) \tag{27-2}$$

and, therefore, the original current without the ammeter is

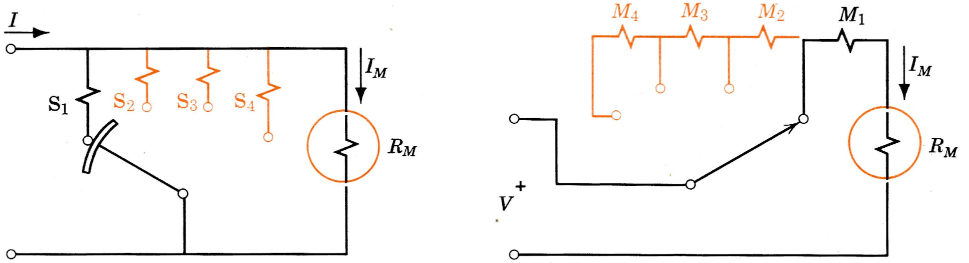
$$I = I_M\left(\frac{R'}{R' - R_M}\right) \tag{27-3}$$

In this case $R' = 0.2 + 2 + (2 \times 2)/(2 + 2) = 3.2\ \text{k}\Omega$ and

$$I = 0.937\left(\frac{3.2}{3.2 - 0.2}\right) = 1\ \text{mA}$$

---

† See p. 373 of H. H. Skilling, *Electrical Engineering Circuits*, 2nd ed., John Wiley & Sons, New York, 1965.



(a) Ammeter with shunts

(b) Voltmeter with multipliers

Figure 27.2 Multirange instrument connections.

inserting a current-limiting multiplier resistance in series with the moving coil. A 25-Ω instrument providing full-scale deflection with a current of 1 mA develops a voltage drop of $25 \times 0.001 = 0.025$ V. To convert this instrument to a 0- to 5-V voltmeter, the multiplier resistance must develop a voltage drop of $5 - 0.025 = 4.975$ V when the meter current is 1 mA. Therefore, $M = 4.975/0.001 = 4975\ \Omega$.

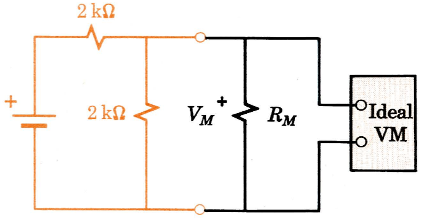
A circuit for a multirange voltmeter is shown in Fig. 27.2b. Four multiplier resistances are available to provide four voltmeter ranges. The values can be calculated from the equations

$$\frac{V_1}{I_M} = R_M + M_1 \qquad \frac{V_2}{I_M} = R_M + M_1 + M_2 \tag{27-4}$$

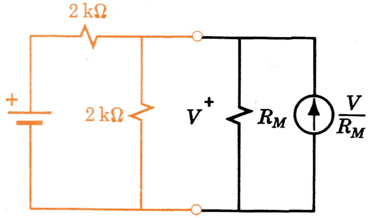
and so forth. Initial voltage readings should always be made on the highest voltage range; a lower range can then be selected to obtain an easily read deflection. In a given measurement, the effect of connecting the voltmeter must be taken into account.

Example 2

A voltmeter with a 1-mA movement ($R_M = 5000\ \Omega$ on the 5-V scale) reads 2.5 V in the circuit of Fig. 27.3a. Predict the voltage without the meter.



(a) Connection



(b) Compensation

Figure 27.3 Effect of voltmeter on voltage.

Following the same line of reasoning as in Example 1, the compensation theorem indicates that the effect of the voltmeter can be compensated by the insertion of a current source V/R_M as shown in Fig. 27.3a. The actual voltmeter reading is

$$V_M = V\left(1 - \frac{R'}{R_M}\right) = V\left(\frac{R_M - R'}{R_M}\right) \tag{27-5}$$

where R' is the resistance seen by the ideal voltmeter. Therefore, the original voltage without the voltmeter is

$$V = V_M\left(\frac{R_M}{R_M - R'}\right) \tag{27-6}$$

In this case, R' is the equivalent resistance of the parallel combination of 2 kΩ, 2 kΩ, and 5 kΩ or $R' = 0.833$ kΩ and

$$V = 2.5\left(\frac{5}{5 - 0.833}\right) = 3\text{ V}$$

In Example 2 the connection of the voltmeter introduces an error of greater than 16%. To keep this error less than 1%, Eq. 27-5 indicates that $R_M/R' \geq 100$ or $R_M \geq 100R' = 83{,}300\ \Omega$. For electronic circuit testing a typical instrument has a 50 μA movement and an internal resistance of 2000 Ω. When the proper multiplier is used to convert this instrument to a 5-V voltmeter, the effective meter resistance is $\frac{5}{50} \times 10^6$ or 100,000 Ω. When converted to a 1-mA milliammeter, the effective meter resistance is 2000 (50 μA/1 mA) = 100 Ω.

The Ohmmeter. A useful application of the d'Arsonval movement is in measuring the dc resistance of passive circuits. A simple ohmmeter is shown in Fig. 27.4. The battery (usually 1.5 V), a fixed resistance R_1, and a variable "Ohms Adjust" resistor R_A are mounted within the ohmmeter case. With the test prods shorted, R_A is varied until the voltmeter reads full scale (usually 1 V). When the test prods are across an unknown resistance R_X, the voltage read is the voltage across R_1. Assuming R_M is very large compared to R_1,

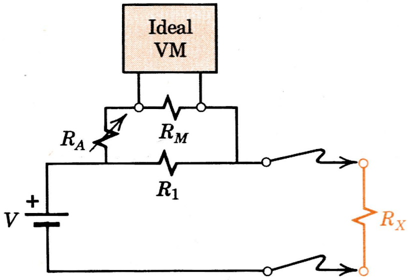


Figure 27.4 A simple ohmmeter.

$$V_M = \frac{R_1}{R_1 + R_X} \times \text{full scale.} \qquad (27\text{-}7)$$

Full-scale deflection corresponds to 0 resistance; half-scale deflection is obtained when $R_X = R_1$. Very high resistance values are crowded into the lower part of the scale. Since the voltmeter reading changes most rapidly with a given percentage change in resistance when $R_X = R_1$ (see Exercise 6), midscale readings are most precise. A selector switch is used to change scales by selecting values of R_1; typical ranges are for midscale readings of 10 Ω, 1 kΩ, and 100 kΩ. Practical ohmmeters use a more complicated circuit† but the general principle of resistance measurement is the same.

Two precautions must be observed in using an ohmmeter. First, resistance measurements can be made only on passive circuits; active sources would contribute currents that would upset the resistance–current relation and perhaps damage the instrument. Second, the ohmmeter should not be used to measure the resistance of elements that would be damaged by the current employed; some semiconductor devices, meter movements, and fuses would be destroyed by a current of 1 mA or so.

---

† See Malmstadt, Enke, and Toren, *Electronics for Scientists,* W. A. Benjamin, Inc., New York, 1963.

The Multimeter. Since the same d'Arsonval movement can serve for current, voltage, and resistance measurements, an effective arrangement is to include shunts, multipliers, and battery in the same case along with the necessary selector switch. Such a combination volt-ohm-milliammeter (VOM) is called a *multimeter.* In addition to dc quantities, ac voltages can be measured by using a rectifier circuit. The precautions noted in connection with the individual instruments apply to the multimeter as well. When the multimeter is not in use, the selector switch should be turned to a high dc voltage scale to avoid battery drain due to accidental short-circuiting of the terminals or damage of the rectifiers or movement by accidental connection to a dc source.

Instrument Characteristics and Errors

The preceding discussion of dc measurements provides a basis for a general description of instrument characteristics and measurement errors. The difference between the measured value and the true value of a quantity is called the *error.* Relatively constant errors that appear in all measurements in a series are called *systematic* errors; these may be due to such effects as miscalibration, nonlinearities, zero shift, or observer bias. Variable or accidental errors that appear in certain measurements are called *random* errors; these may be due to such effects as bearing stickiness, observer mistake, or to unknown causes.

No physical measurement is without error. The *accuracy* of a measurement is the ratio of the error in the measured value to the true value (usually defined by a more accurate measurement that in turn includes some error) and is usually expressed in percent. In producing a voltmeter with a specified accuracy of 0.5% of full scale, the manufacturer employs designs, fabrication processes, and testing procedures that ensure that a certain fraction, say 98%, of all voltmeters produced are within the specified accuracy. The term *precision* is defined as a measure of the reproducibility of independent measurements by the same instrument and is not to be confused with accuracy.†

In making any measurement there is always some effect on the quantity being measured. For an accurate measurement of an electrical quantity there are two possible approaches. One is to make the measurement and then correct for the disturbance due to measuring. The other is to make the measurement in such a way that the effect on the measured quantity is less than a certain acceptable amount.

Examples 1 and 2 illustrate these approaches. Using the compensation theorem, the effect of connecting an ammeter or voltmeter can be estimated and the indicated current or voltage can be corrected. This approach has wide application throughout engineering and science. Knowing the parameters of the circuit and the effect of connecting an ammeter or voltmeter, we can specify the characteristics of the measuring instrument. From the examples we conclude that for negligible error an ammeter should have a low resistance compared to that seen by an ideal ammeter inserted at the same point and a voltmeter should have a high resistance compared to that seen by an ideal voltmeter connected to the same terminals.

---

† For a more complete discussion see E. Frank, *Electrical Measurement Analysis,* McGraw-Hill Book Co., New York, 1959.

There are other instrument characteristics that require attention but can only be mentioned here. Since the *response* of an instrument is never instantaneous, the indicated value is never the true value when a variable is changing. The time *lag* is dependent on the natural frequency of the instrument and the degree of damping. Some instruments are subject to *hysteresis;* the response to a given signal depends on the previous condition. Also there may be a slow change in properties with time, called *drift,* which influences the calibration. Freedom from the influence of environmental conditions such as temperature or humidity is referred to as *stability.*

AC Measurements

In the *iron-vane* instrument of Fig. 23.20a, the developed torque is proportional to the square of the current and therefore alternating currents can be measured. In the *thermocouple* instrument, a d'Arsonval movement is used to measure the emf generated at a junction of dissimilar materials. The current to be measured flows through a heater element and heats the thermocouple junction. Since the dc emf is dependent on the temperature rise, which is nearly proportional to the square of the ac current, the instrument can be calibrated in terms of rms currents. The thermocouple meter can be used to measure true rms values at frequencies up to 50 MHz.

Electrodynamometer Instruments. In the *electrodynamometer* instrument of Fig. 3.22b, the developed torque is proportional to the product of the current in a fixed coil and the current in the moving coil. In an ammeter the same current flows in both coils and the torque is proportional to the square of the current; the inertia of the movement performs an averaging function and the indication is a true rms value. In a wattmeter the current in one coil is proportional to the instantaneous current and the current in the other coil is proportional to the instantaneous voltage; the torque is proportional to the instantaneous power and the indication is proportional to the average power. Electrodynamometer instruments are useful for accurate measurements at frequencies up to a few hundred cycles per second.

Rectifier Instruments. Semiconductor diodes can be used to convert alternating current to direct current, which is measured by a d'Arsonval movement. In the full-wave bridge circuit (Fig. 27.5), an alternating current at the terminals produces a unidirectional current through the moving coil and the deflection is proportional to the average current. These instruments are calibrated on sinusoidal waves to read rms values. Since the rms value of the sinusoid is $0.707/0.636 = 1.11$ times the average value, the instrument indicates 1.11 times the average value of the rectified waveform of any applied current. It is accurate for sinusoids only; for any other waveform a correction must be applied.

Example 3

A rectifier-type ammeter indicates a current of "2.22 A rms" when measuring a triangular wave (Fig. 27.5). Estimate the peak and rms value of the current.

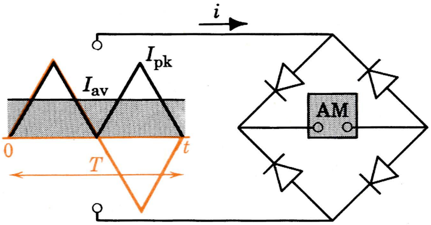


Figure 27.5 A rectifier-type instrument.

This instrument indicates 1.11 times the average of the rectified wave; therefore, the average is

$$I_{av} = \frac{2.22}{1.11} = 2 \text{ A}$$

and the peak value for the triangular wave is

$$I_{pk} = 2I_{av} = 4 \text{ A}$$

The true rms value is

$$I_{rms} = \sqrt{\frac{1}{T}\int_0^T i^2\,dt} = \sqrt{\frac{I_{pk}^2}{3}} = \sqrt{\frac{16}{3}} = 2.31 \text{ A}$$

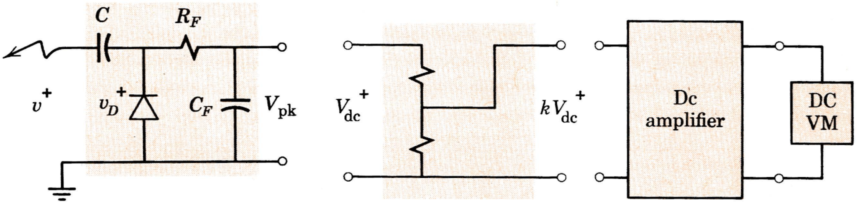
Conclusion: For triangular waves the reading is about 4% low.

Electronic Indicating Instruments

The availability of stable, high-gain electronic amplifiers whose characteristics are known precisely has contributed greatly to the art of instrumentation. One reason for using amplifiers is to reduce the power required from the system under observation; the power required to drive the indicating or recording instrument is provided by the amplifier. A related advantage is that the input impedance of an electronic voltmeter, for example, can be millions of ohms instead of the few thousand ohms of a d'Arsonval instrument. Another reason is that electronic devices can follow signal variations at frequencies of many megacycles per second. Amplifiers are particularly valuable when the signal being measured is very small and might be lost in the *noise* generated by extraneous sources; a high *signal-to-noise ratio* is essential for accurate instrumentation or precise control.

The Electronic Voltmeter

By using an FET or a vacuum tube as an amplifier, a voltage can be measured with negligible loading effect. Because of the importance of the electronic voltmeter (EVM), highly refined circuits have been developed to provide accuracy, flexibility, and stability. In the voltmeter of Fig. 27.6bc, the voltage to be



(a) An ac input circuit (b) Attenuator (c) Amplifier and voltmeter

Figure 27.6 Basic components of an electronic voltmeter.

measured is attenuated to the proper level by a calibrated potentiometer and applied to the FET input stage of a balanced dc amplifier (similar to the differential amplifier of Fig. 19.21a with the second input grounded so that $v_n = 0$.) The gate current is very small so that the input impedance can be many megohms. The direct-coupled differential amplifier is stabilized by feedback and has a low impedance output (emitter follower) to drive the d'Arsonval movement.

To measure alternating voltages, a diode clamper is used (see explanation accompanying Fig. 12.25). In Fig. 27.6a, capacitor C tends to charge up to the negative peak value V_{pk} of the input voltage. The diode voltage v_D is then equal to $v + V_{pk}$ or the clamper output contains a dc component equal to the peak value of the input. The combination of R_F and C_F is a filter that removes the ac component, and dc voltage V_{dc} is applied to the amplifier. This type of EVM responds to peak values (of symmetrical waves) but is calibrated to read rms values of sinusoids; in other words, the meter indication is 0.707 times the positive peak value of the input. On other than sinusoids, a correction must be applied.

Example 4

The triangular current wave of Example 3 indicated "2.22 A rms" on a rectifier-type ammeter. Predict the indication on a peak-responding EVM connected across a 1-Ω resistor carrying this current.

The peak value of the triangular voltage wave is

$$I_{pk}R = 4 \times 1 = 4\text{ V}$$

and the EVM indicates 0.707 times the peak value or

$$V_M = 0.707 \times V_{pk} = 0.707 \times 4 = 2.83\text{ V}$$

The rms value is 2.31 V (see Example 3); therefore, for triangular waves the indication is approximately 22% high.

On a commercial EVM, a switch converts the instrument from dc to ac operation. The input impedance is 10 MΩ or more and remains constant up to frequencies of 10 MHz or greater. Two clamping circuits are used to obtain true peak-to-peak values for asymmetric waves. For measuring small signals at high frequencies, the input diode is placed in the test probe itself and the effect of the capacitance of the leads is minimized.

For measuring small ac currents, a special probe clamps around the current-carrying wire and develops a voltage proportional to the current encircled. For measuring resistance, a standard dc current is sent through the unknown resistance and the voltage drop is indicated on a scale calibrated in ohms.

The Electronic Counter

A recent instrumentation development of great value is the electronic counter diagramed in Fig. 27.7. The events to be counted—meteorites, heartbeats, or revolutions—are transduced into electrical signals that are then transformed into sharp pulses. A precision time base is established, perhaps by a controlled oscillator, and used to open and close the gate to the cascaded scale-of-ten counters for display. The counters consist of four flip-flops (see Fig. 14.20 and accompanying discussion) with feedback loops to convert from a scale of sixteen to a decimal scale. In counting, the gate is held open for a precise fraction of a second, the events are counted, and then the gate is closed and the total count displayed.

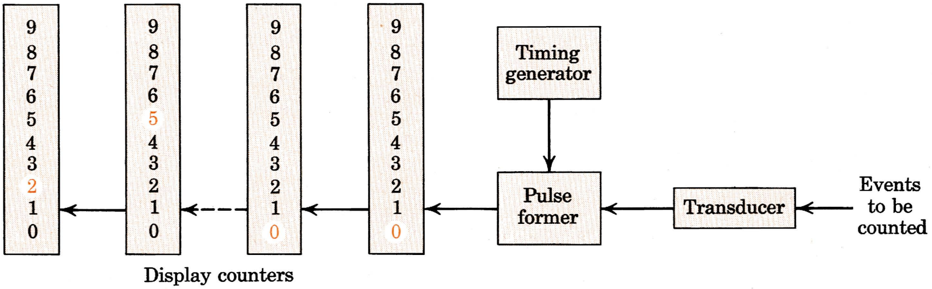


Figure 27.7 Elements of an electronic counting system.

The same unit can be adapted to measure the *elapsed time* between similar events. In this mode of operation, an event is used to open the gate and the cycles of the timing generator are counted until the next event closes the gate. In this manner the period of an electrical signal can be counted by determining the time between corresponding points on the wave. If the timing generator operates at a frequency of 10 MHz (accurate to ±1 Hz), and if 2500 pulses are counted during a given time interval, the time interval is 0.0002500 s. If this represents the period of a sine wave, the frequency is 4000 Hz and the counter is performing *frequency measurement.*

The Digital Voltmeter

The DVM offers many advantages over other voltmeters: greater speed, higher accuracy and resolution, reduced operator error, and compatibility with other digital equipment. The widely used *dual-slope technique* is illustrated in Fig. 27.8. The unknown input voltage V_X is applied to the integrator at time t_1, and capacitor C charges at a rate proportional to V_X for a fixed time T_F. At time t_2 (after a fixed number of clock pulses has been counted), the control logic switches the integrator input to fixed reference voltage V_F and C discharges at a rate proportional to V_F. The counter is reset at t_2 and counts until the comparator indicates that the integrator output voltage v_P has returned to zero at t_3. By similar triangles,

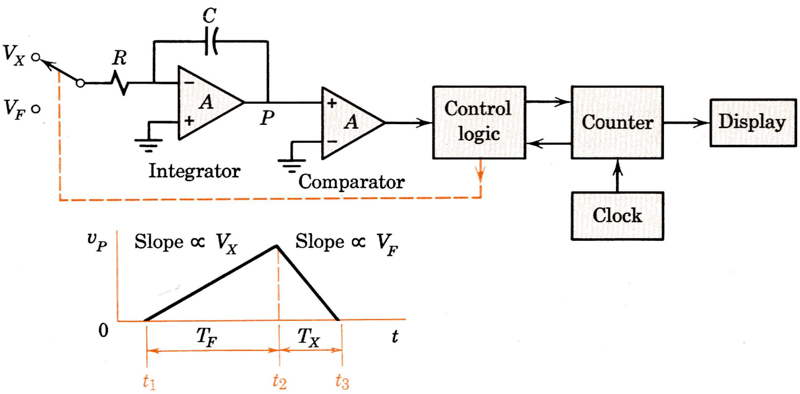


Figure 27.8 A digital voltmeter using dual-slope integration.

$$V_X T_F = V_F T_X \qquad \text{or} \qquad V_X = V_F(T_X/T_F)$$

The count at t_3, proportional to the input voltage, is displayed as the measured voltage.

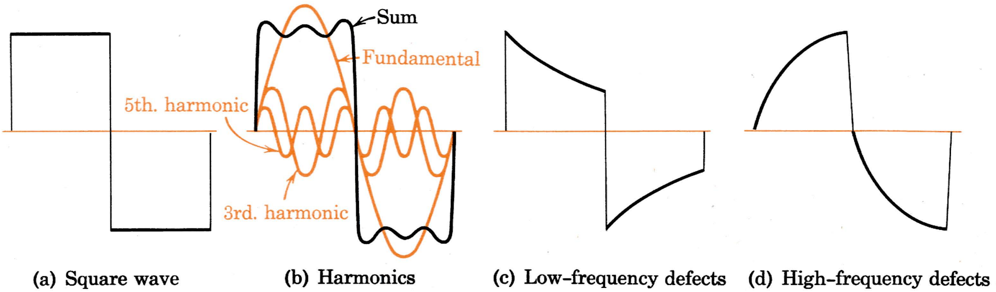
The integration process averages out noise, and by integrating over 1/60 s, 60-Hz hum is almost eliminated. Three to five readings are taken per second. By using the appropriate *signal conditioner,* currents, resistances, or ac voltages can be measured by the same instrument. Such a digital multimeter (DMM) may incorporate automatic range and polarity selection so that measurement is reduced to selecting a variable and pressing a button.

The Cathode-Ray Oscilloscope

The CRO is a versatile instrument with nearly ideal input characteristics and excellent frequency response (Fig. 10.8). Basically, it provides a spot deflection directly proportional to an instantaneous voltage. In its practical form, it is the most flexible laboratory instrument with controls conveniently arranged for precise measurement over a wide range of test conditions—from microvolts to hundreds of volts and from dc to many megahertz.

The basic applications of the CRO to *x-y* plotting, frequency comparison, and the measurement of voltage, time, and phase are described in Chapter 10. In the most common laboratory use, an input voltage signal is applied to the vertical amplifier and a calibrated internal sweep generator is used to provide a time axis for *waveform display*. This arrangement permits the study and analysis of any physical variable that can be converted into a voltage, for example, the pressure changes in an engine cylinder during combustion or the stresses in a building column under earthquake forces.

The unique convenience of a CRO is illustrated in the measurement technique called *square-wave testing*. Any periodic wave can be represented by a Fourier series of sinusoids. As indicated in Fig. 27.9b, the sum of the first three odd harmonics (with appropriate amplitude and phase) begins to approximate a square wave; additional higher harmonics would increase the slope of the leading edge and smooth off the top. If a square-wave input to a device under test results in an output resembling Fig. 27.9c, we conclude that low-frequency components have been attenuated and shifted forward in phase. If the output resembles Fig. 27.9d, we conclude that high-frequency components have been attenuated and shifted backward in phase. The frequency response of an amplifier, for example, can be easily determined by varying the frequency of the square-wave input until these distortions appear (near half-power frequencies f_1 and f_2). As another example, two devices giving similar responses to a square-wave input can be expected to respond similarly to other waveforms.



(a) Square wave (b) Harmonics (c) Low-frequency defects (d) High-frequency defects

Figure 27.9 Square-wave testing with a CRO.

Every year sees new advances in CRO design as the instrument manufacturers strive to meet new needs of the laboratory and the field by taking advantage of newly developed devices and techniques. Among the features offered by modern CROs are:

Differential Inputs. Each amplifier channel has two terminals beside the ground terminal. By amplifying only the difference between two signals, any common signal such as hum is rejected.

Optional Probes. To improve the input characteristics of a CRO or to expand its functions, the input stage can be built into a small unit (connected to the CRO by a shielded cable) that can be placed at the point of measurement. *Voltage probes* insert impedances in series with the CRO input to increase the effective input impedance. *Current probes* use transformer action or the Hall effect to convert a current into a proportional voltage for measurement or display.

Dual Channels. The waveforms of two different signals can be displayed simultaneously by connecting the signals alternately to the vertical deflection system. The more expensive *dual beam* feature requires separate electron guns and deflection systems but permits greater display flexibility.

Controlled Triggering. The instant on the triggering signal at which a sweep begins is controlled by the *trigger slope* switch (positive or negative) and the *trigger level* setting (sign and magnitude). For example, if the *slope* is set to "+" and the *level* is set to "0," the sweep will begin when the triggering signal goes through zero with a positive slope.

Delayed and Expanded Sweep. To display a magnified version of a selected, small portion of a waveform, an auxiliary delayed sweep with a faster sweep speed can be used.

Storage. If the input signal is a single, nonrepetitive event, the image can be *stored* by using a long persistence phosphor. By placing a storage mesh directly behind the phosphor and controlling the rate at which the charge pattern leaks off the mesh, a *variable persistence* is obtained.

Sampling. To display signals at frequencies beyond the limits of the CRO components, very short *sample* readings are taken on successive recurrences of the waveform. Each amplitude sample is taken at a slightly later instant on the waveform and the resulting dots appear as a continuous display. With this sampling technique, 18-GHz signals can be displayed.

Digital Readout. The addition of a built-in LSI *microprocessor* provides direct digital readout of time interval, frequency, or voltage in addition to the conventional CRO display. The operator sets two markers to indicate a horizontal or vertical displacement on the waveform; the microprocessor interrogates the function switches and the scale switches, calculates the desired variable, and converts it to digital form for LED display.

The price paid for the extreme flexibility of a modern CRO when used properly is the likelihood of spurious or confusing displays when the controls are improperly set. A careful study of the instruction manual and a step-by-step run-through of the suggested procedures are essential when getting acquainted with a new instrument.

Instrumentation Transducers

The convenience of the electronic voltmeter, the versatility of the cathode-ray oscilloscope, and the speed and precision of the electronic counter are available for measuring any physical variable that can be transduced to electrical form. The *transducers* or *sensors* employ a great variety of physical-electric phenomena, but they can be classified into two general categories.

Active Transducers

A sensing element which generates an emf or current is called an *active transducer*. The electrical power generated may be sufficient to drive a d'Arsonval movement, as in the case of a thermocouple, or it may be so minute that special high-impedance amplifiers must be used (see p. 515), as in the case of the pH probe. Useful active transducers include the following.

Thermoelectric. The emf developed across a junction of two dissimilar materials is a sensitive function of temperature. If a circuit contains two junctions at different temperatures, a net emf is developed. In practice, the reference junction is maintained at a known temperature and the other junction is exposed to the temperature to be measured (see Fig. 27.10). For commercial *thermocouples*, tables are available for direct conversion of the observed emf to the corresponding temperature difference. Thermocouples are rugged and accurate, but the fairly long response time constant prevents accurate measurement of rapidly changing temperatures.

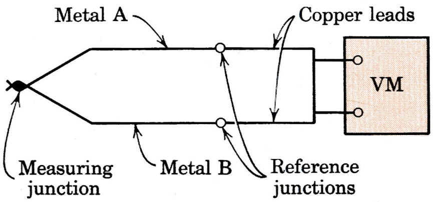


Figure 27.10 A thermocouple circuit.

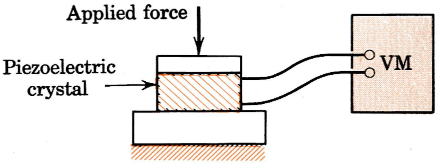


Figure 27.11 A piezoelectric transducer.

Piezoelectric. Certain crystalline materials, natural and synthetic, develop an emf when they are deformed. This *piezoelectric* effect is used in microphones and phonograph pickups and also in instruments measuring displacement or pressure (see Fig. 27.11). The small size and very good high-frequency response of crystal elements make them valuable in dynamic measurements.

Photoelectric. The sensitivity of cathode surfaces and semiconductor junctions to illumination permits the use of *phototubes* and *photojunction cells* as optical transducers. In the phototube, incident photons cause the emission of electrons with sufficient energy to move to the anode and to contribute to an external current. In the photodiode, high-energy photons create electron-hole pairs in the depletion region and contribute to an external current just as in the solar cell.

Electromagnetic. The moving coil in a magnetic field employed in a dynamic microphone or a dynamic pickup is a velocity-sensitive transducer. Linear displacement and acceleration measurements can be obtained by electronically integrating and differentiating the output of a velocity transducer.

Angular velocity can be measured by using dc or ac *tachometer* generators with permanent magnets; the generated emf is directly proportional to the velocity. The output of an ac tachometer is usually rectified and measured on a dc instrument. Another method of measuring angular velocity is shown in Fig. 27.12. The projecting teeth vary the reluctance in the air gap of the permanent magnet, inducing pulses that are counted by an electronic counter.

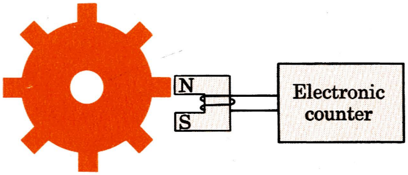


Figure 27.12 An electromagnetic revolution counter.

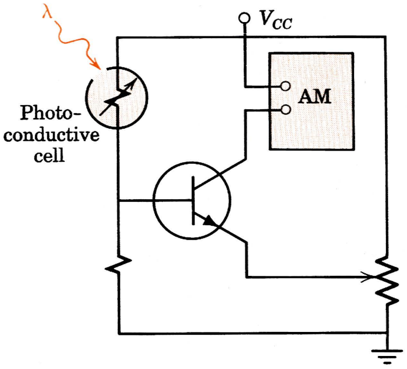


Figure 27.13 A photoconductive cell circuit.

Passive Sensors

Any physical effect that changes an electrical circuit parameter can be used in sensing. In such *passive sensors* a source of energy must be available.

Resistance. The resistance of a conductor may be affected by temperature, mechanical strain, or illumination. The *resistance thermometer* is widely used for precise continuous and remote temperature measurement. The *hot-wire anemometer* uses a variable resistance to measure the velocity of the air stream in which it is immersed. A fine wire is heated by an electric current and the voltage drop observed; the voltage and the resistance are dependent on wire temperature that is governed by the cooling effect of the moving air. In the *strain gauge,* stretching a fine wire increases its length, reduces its cross section, and changes its conductivity; the resulting change in resistance is approximately proportional to the strain. In the *photoconductive cell* an increase in illumination increases the rate of generation of electron-hole pairs and increases the conductivity. In the circuit of Fig. 27.13, the voltage-divider is adjusted until the transistor is cut off with the photoconductive cell dark; illumination results in a change in emitter-base bias and a large increase in collector current.

Inductance. The common *variable reluctance* sensor uses the displacement of a magnetic armature to change the reluctance and thereby alter the self-inductance of a coil. Any quantity such as force, pressure, acceleration, or velocity that can be converted to a displacement can be measured indirectly. A unique application is the measurement of the thickness of paint on magnetic materials; when the U-shaped core is placed on a flat surface, the paint forms two gaps in the magnetic circuit. In the important *differential transformer* (discussed in the following section), mutual inductance is varied by the displacement of a magnetic core linking two coils.

Capacitance. The capacitance of a parallel-plate unit is dependent on the plate separation and the dielectric material. In the *capacitance pressure* sensor (Fig. 27.14a), a variation in pressure flexes a conducting diaphragm and decreases its separation from a fixed plate. In the *liquid level gauge* (Fig. 27.14b), the liquid has access to the space between two parallel plates; since the dielectric constant of the liquid differs from that of air, the capacitance of the unit is a function of liquid height. The variation in capacitance can be used to change the response of a parallel circuit as shown in Fig. 27.14c or change the frequency of a resonant circuit.

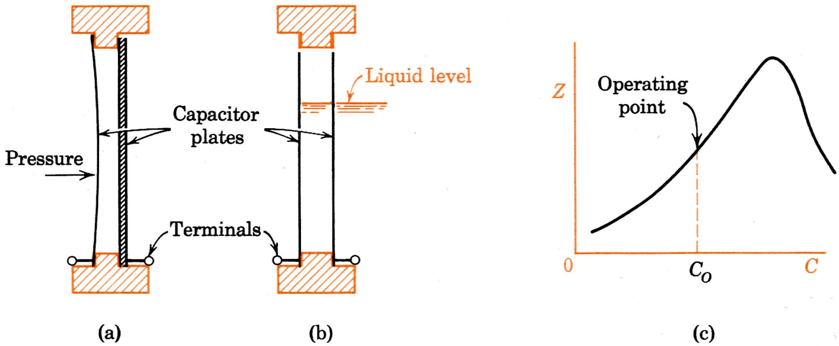


Figure 27.14 Variable-capacitance transducers and characteristic.

Bridge Measurements

With a passive sensor, an external electrical source is required. One possibility with a resistance thermometer, for example, is to use a dc supply and measure the change in current resulting from the change in resistance (as in the ohmmeter). But when the expected change in current is small, even a good ammeter cannot provide an accurate measurement. For accurate measurement of passive elements, the *bridge* circuit is an excellent arrangement.

The Wheatstone Bridge. In the basic Wheatstone bridge shown in Fig. 27.15, R_A, R_B, and R are standard resistances, and R_X is the resistance to be measured. The *null detector* D is a sensitive galvanometer that reads zero when the bridge is *balanced.* At balance there is no current from a to b; therefore, nodes a and b are at the same potential and

$$\begin{aligned} I_A R_A &= I_B R_B \\ I_A R_X &= I_B R \end{aligned} \qquad (27\text{-}8)$$

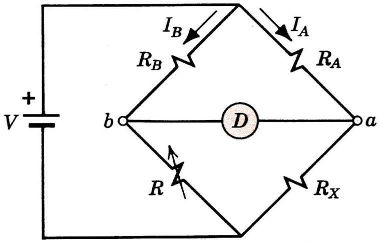


Figure 27.15 Wheatstone bridge.

Dividing the first equation by the second and solving,

$$R_X = R\frac{R_A}{R_B} \qquad (27\text{-}9)$$

Since the three standard resistances are known within about 0.01%, if the null point is established precisely, the unknown resistance can be measured with a high degree of accuracy. Modifications of this basic bridge circuit provide for measurement of large and small resistances and of small changes in resistance such as those encountered in strain gauge measurements. Note that balance is independent of the applied voltage; a stable, precisely known voltage is not required. If an ac source is used, the detector can be an earphone. The combination of an IC op amp and an earphone can provide the same sensitivity as an expensive galvanometer.

Example 5

An unknown resistance is measured on a Wheatstone bridge and a null is obtained with $R_A = 500\ \Omega$, $R_B = 5000\ \Omega$, and $R = 270.4\ \Omega$. The fixed resistors are accurate within ±0.01% and the variable resistor within ±0.02%. Determine the unknown resistance.

Assuming the null point is precisely determined, from Eq. 27-9,

$$R_X = R\frac{R_A}{R_B} = 270.4(1 \pm 0.0002)\frac{500(1 \pm 0.0001)}{5000(1 \pm 0.0001)}$$

The worst case is that in which R and R_A are in error in one direction and R_B in the other. In this case the errors add and

$$R_X = 27.04(1 \pm 0.0004) = 27.04 \pm 0.01\ \Omega$$

In computing errors, as in Example 5, two approximate formulas are useful. For x, y, and z very small compared to 1,

$$(1 + x)(1 + y) = 1 + x + y + xy \cong 1 + x + y$$

and (27-10)

$$\frac{1}{1 + z} = (1 + z)^{-1} = 1 - z + z^2 - \cdots \cong 1 - z$$

The Impedance Bridge. The same bridge circuit can be arranged to measure impedances as shown in Fig. 27.16. Here a null is obtained when

$$\mathbf{Z}_X = \mathbf{Z}\frac{R_A}{R_B} \tag{27-11}$$

Since $\mathbf{Z}_X = R_X - j(1/\omega C_X)$, Eq. 27-11 is satisfied only if

$$R_X = R\frac{R_A}{R_B} \quad \text{and} \quad C_X = C\frac{R_B}{R_A} \tag{27-12}$$

In words, the real and imaginary components of impedance must balance separately.

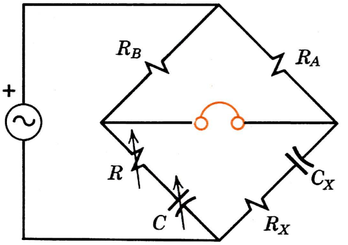


Figure 27.16 Impedance bridge.

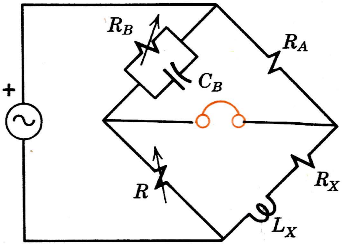


Figure 27.17 Maxwell bridge.

In deriving Eqs. 27-12 it is assumed that the unknown consists of a series combination of R_X and C_X. Various forms of the basic bridge have been invented to measure unknown impedances accurately and conveniently. In the *Maxwell bridge* of Fig. 27.17, a standard capacitor C_B permits the measurement of an unknown inductive impedance. It can be shown (see Exercise 17) that the balance equations for this circuit are

$$R_X = R\frac{R_A}{R_B} \quad \text{and} \quad L_X = C_B R R_A \tag{27-13}$$

This is convenient because standard fixed capacitors are less expensive than standard variable inductors.

Dynamic Bridge Response. The null method works well in making static measurements, but it cannot be used in measuring rapidly changing parameters.

In determining the dynamic stresses in a structure, using strain gauges for example, a practical method is to balance a Wheatstone bridge under static conditions and then use the unbalanced bridge as a dynamic transducer. A convenient approach is to represent the bridge by a Thévenin equivalent circuit.

For the special case of $R_A = R_B = R$, the Wheatstone bridge is as shown in Fig. 27.18a. To determine the signal voltage v due to a very small resistance change ΔR, we note that

$$V_1 = \frac{R}{2R} V = \frac{V}{2}$$

and

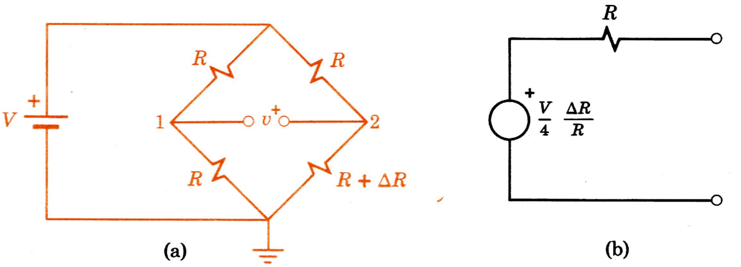
$$V_2 = \frac{R + \Delta R}{R + R + \Delta R} V = \frac{V}{2} \frac{1 + \Delta R/R}{1 + \Delta R/2R} = \frac{V}{2} \frac{1 + \delta}{1 + \delta/2}$$

where $\delta = \Delta R/R$, the fractional unbalance. Using Eqs. 27-10,

$$V_2 \cong \frac{V}{2}(1 + \delta)\left(1 - \frac{\delta}{2}\right) \cong \frac{V}{2}\left(1 + \frac{\delta}{2}\right)$$

and

$$-v = V_2 - V_1 \cong \frac{V}{2}\left(1 + \frac{\delta}{2}\right) - \frac{V}{2} = V \cdot \frac{\delta}{4} = \frac{V}{4} \frac{\Delta R}{R} \tag{27-14}$$



(a) (b)

Figure 27.18 An unbalanced bridge and the Thévenin equivalent.

The resistance looking into the bridge with the source voltage removed is just $2(R/2) = R$ and the Thévenin equivalent is as shown in Fig. 27.18b. The output signal is small, but it is easily amplified and recorded or displayed on an oscilloscope. Another possibility is to excite the bridge with a high-frequency ac source; in this case the output is an amplitude-modulated wave particularly well suited to telemetering. The dynamic stresses in a manned spacecraft during take-off can be measured in this way.

Potentiometer Measurements

In bridge measurements an unknown impedance is compared to a standard. The same approach is employed in the analytical balance where an unknown weight is compared to a standard weight under conditions of balance. Such *comparison methods* are capable of great accuracy and are widely used in science and engineering. The *potentiometer*† is a useful instrument for measuring voltages by comparison and is frequently used in conjunction with active transducers.

The Slide-Wire Potentiometer. If an accurately known voltage is connected across a linear resistance, such as a uniform wire, a variable standard voltage is available for comparison with an unknown. One practical form of this arrangement is shown in Fig. 27.19. A source V supplies a standard current I_S through a uniform potentiometer resistance R_P, an adjustable resistance R_A, and a standard resistance R_S. A *standard cell* provides a precisely known emf V_S when no current is drawn. To set the current to the standard value, switch S is placed in the position shown, and R_A is adjusted until detector D indicates that $I_S R_S$ is equal to the emf of the standard cell.

With $I_S = V_S/R_S$ accurately known, the potentiometer voltage $V_O = I_S R_O$ is determined solely by the position of the sliding contact. To measure an unknown voltage, switch S is thrown and the sliding contact is moved until the detector indicates a null. In practical form, R_P is a compact multiturn resistance with a continuously variable tap; the dial is calibrated to read voltage directly. For greater accuracy a series combination of fixed and variable resistances is used for R_P. For greater convenience the voltage source V can be replaced by a Zener-diode–regulated voltage supply and the standard cell omitted.

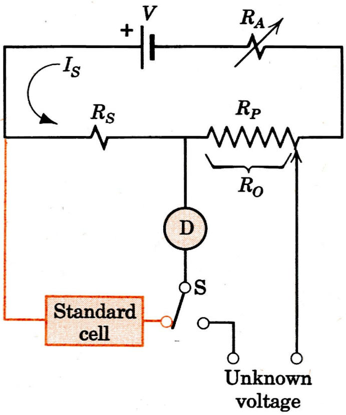


Figure 27.19 A slidewire potentiometer.

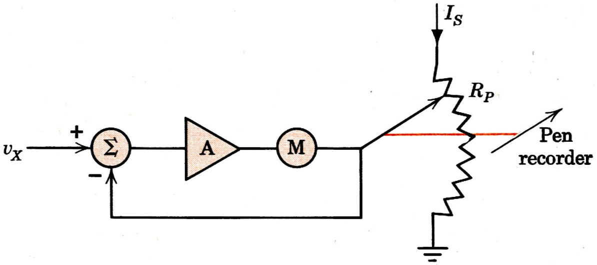


Figure 27.20 A self-balancing potentiometer.

Self-Balancing Potentiometer. For automatic control and continuous recording, the self-balancing potentiometer is widely used. In the simplified block diagram of Fig. 27.20, the difference between the input voltage and the potentiometer voltage is amplified and used to drive the balancing motor M. Rotation of the motor shaft is always in such a direction as to decrease the unbalance between the potentiometer voltage and the voltage being measured. When the system is in balance, the position of the shaft indicates the input voltage (from a thermocouple, for example) and the pen provides a record (of temperature) on a strip chart. In one type of X-Y recorder, self-balancing potentiometers provide vertical and horizontal deflections proportional to input voltages corresponding to the two variables being plotted on the stationary paper.

---

† Here *potentiometer* refers to a measuring instrument employing a voltage divider; the voltage divider itself is also called a potentiometer.

The Strain-Gauge Transducer

Two passive transducers are used so widely that they deserve further discussion. One is the *bonded resistance strain gauge,* consisting of a grid of fine resistance wire cemented or bonded to a thin paper backing (Fig. 27.21). When the gauge is cemented to a structural member under test, any deformation of the member results in a change in dimensions and therefore a change in resistance of the gauge. The relation between strain ϵ and resistance change is

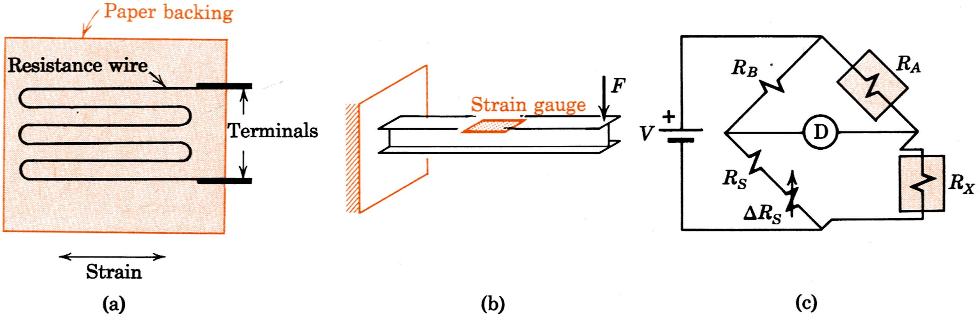


Figure 27.21 Application of a bonded resistance strain gauge.

$$\frac{\Delta R}{R} = K_G\epsilon = K_G\frac{\Delta l}{l} \tag{27-15}$$

where K_G is the gauge factor. For most metallic elements the gauge factor is between 2 and 4; semiconductor strain gauges are available with gauge factors of several hundred.

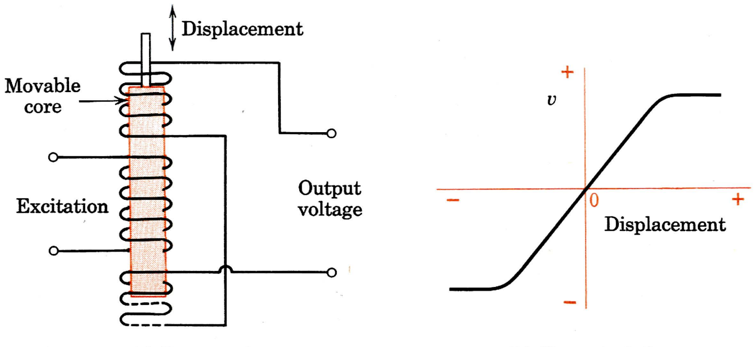
Since the strain $\Delta l/l$ is usually very small, the change in resistance is small and bridge methods are indicated. In the arrangement shown in Fig. 27.21c, R_X is the resistance of the gauge. If the bridge is initially balanced with standard resistance R_S, the change in gauge resistance is measured accurately by the resistance ΔR_S needed for rebalancing after strain occurs.

One problem is that the bridge cannot distinguish between changes in resistance due to strain and those due to temperature changes. The effect of temperature can be reduced by using materials having very low temperature coefficients of resistance. One clever arrangement is to use for R_A an unstressed dummy gauge exposed to the same temperature as R_X. Another possibility is to mount the second gauge symmetrically with respect to the first so that when one is in tension the other is in compression. The effects of temperature and gauge lead wire are canceled out and the sensitivity of the bridge is doubled.

Although the strain gauge is directly responsive to small displacements, it can be used for indirect measurement of other variables such as pressure or acceleration, which can produce strain in an auxiliary mechanism. In wind-tunnel tests of aircraft models, for example, the model is supported by a beam or *sting* with several degrees of freedom. Strain gauges properly placed on the beam permit the determination of lift, drag, and lateral forces and roll, pitch, and yaw torques.

The Differential Transformer

Another very useful device for conversion of mechanical displacements into electrical signals is the *linear differential transformer* shown in Fig. 27.22. The transformer consists of a primary and two identical secondaries connected in series opposition. With the freely moving iron core in dead center, the voltages induced in the two secondaries just cancel and the output is zero. An upward displacement increases the coupling with the upper secondary and decreases the coupling with the lower secondary. A downward displacement produces the opposite effect. The result is an output voltage dependent upon core position.



(a) Construction (b) Characteristic

Figure 27.22 The differential transformer.

The differential transformer is a low-impedance device producing an output of several volts. With proper design the output voltage is a linear function of displacement over a range that may be a few microinches or several inches. This transducer is particularly useful in measuring displacements larger than those for which the strain gauge is suitable. With appropriate auxiliary mechanisms it can be used to measure velocity, acceleration, force, pressure, liquid level, or rate of flow.

Instrumentation Systems

An instrumentation system is an assemblage of components arranged to perform an overall measurement function. But the components must not only perform their individual functions well, they must also work effectively with the associated components. In this sense an instrumentation system is more than the sum of its parts.

System Aspects

In an instrumentation system there is a flow of information in the form of signals. Systems are usually characterized by size and complexity, and in a typical instrumentation system the signals may be transformed many times and into many forms. For example, pressure may affect the position of a diaphragm, whose movement changes the mutual inductance in a differential transformer and modulates the excitation voltage, which controls the frequency of a carrier current, which creates an electromagnetic wave, which is radiated through space. After transmission, reception, and demodulation, the resulting current may deflect a galvanometer, which moves a light beam across photographic paper and leaves a record in the form of a chemical change.

Each component modifies the signal. The signal may deteriorate as a result of time lag or distortion, or the information content may be enhanced by amplification or filtering. For maximum power transfer or optimum response the components must be matched. If the system is to supply data to or receive instructions from a human being, the color of a dial, the shape of a knob, or the pitch of a warning signal may be a critical design element.

Sensors and Transducers

In describing the components of a general data-acquisition system (see Fig. 27.23), let us start with the sensors that detect and respond to the physical variables to be measured. The essential characteristic of a good sensor is that it responds in a predictable and useful way to the quantity being observed; for example, a bimetallic strip bends a known amount in response to a given change

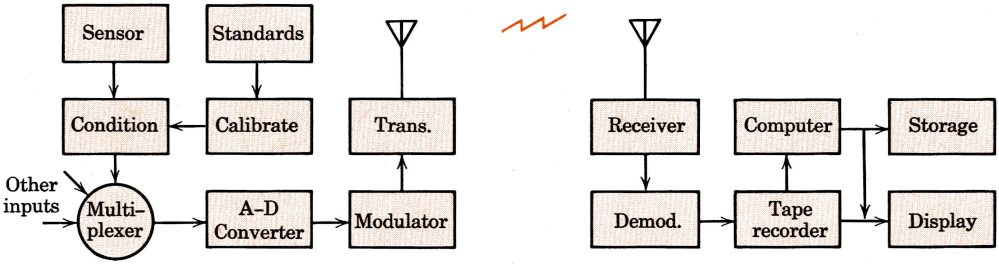


Figure 27.23 A general data-acquisition and reduction system.

in temperature. As mentioned previously, many different physical phenomena are employed in sensors.

Usually we want the response in the form of an electrical signal. The combination of a bimetallic strip and the contacts of an electrical switch provides such an electrical output. In addition to the active transducers previously described, passive sensors such as strain gauges and differential transformers can provide electrical outputs when properly excited. So-called *conditioning equipment* includes the power supplies for bridges, the excitation carriers for differential transformers, the voltage references for potentiometers, and the temperature controlling devices for thermocouples.

Calibrating Equipment

An important part of the "count-down" in a missile firing or any other comprehensive test is the calibration of the sensing elements. Calibration usually involves comparison against a standard. Standards of great precision are available for time, frequency, and voltage. *Atomic clocks* employing the natural resonance frequency of excited atoms keep time with less than 0.01 s deviation per year. The National Bureau of Standards broadcasts over station WWV time signals accurate to 1 μs and audio frequencies accurate to 1 part in 50 million. The primary standard of voltage in this country is a group of 47 Weston cells with a precision of the order of 1 μV.

In the calibration procedure, a series of standard inputs is applied to each channel of the data-acquisition system. The responses are recorded and provide the basis for calibration. Sensing elements that cannot be calibrated in position are periodically calibrated in a testing laboratory and before and after critical measurements.

Analog-to-Digital Converters

Digital signals are less susceptible than analog signals to deterioration by noise or stray pickup or by attenuation in long cables and radio links. Since each binary digit or *bit* of information exists in one of only two states, rather large changes in waveform or amplitude can be tolerated without loss of information. Another reason for converting signals from analog sensors into digital form is to obtain data acceptable to digital computers.

The dual-slope integration used in the DVM is one form of A–D converter. In another method, an analog signal in the form of a varying frequency can be converted into digital form with the aid of an electronic counter. The gate is held open for a fixed interval of time and the number of cycles of the signal frequency fed into a binary counter is a measure of the analog signal.

A direct method of converting analog to digital numbers is the *shaft position* *encoder.* A flat disk commutator rotating with the shaft is divided into annular rings, one for each digit in the number, with a brush riding on each ring. The rings are conducting in the areas for which a logic 1 is to be read. The conducting pattern in contact with the brushes provides the binary code corresponding to the analog position of the commutator. Commercial encoders provide up to 1000 positions per revolution.

Telemetering Equipment

The communication of data from the point of measurement to the point of processing or display is an important element in any instrumentation system. Data transmission to a remote location in the form of electrical signals is called *telemetry* and may be by radio wave or by telephone-type land lines. Land-line telemetry is used in oil refineries, chemical processing plants, electrical power systems, and pipe-line transmission systems. Radio telemetry is used for in-flight tests of aircraft and satellites and for data transmission over great distances or rough terrain.

In radio telemetry the high-frequency carrier wave is modulated by subcarriers upon which the data are superimposed. A great variety of modulation or coding systems is available. To accommodate data from more than one sensor, the capacity of a channel is shared in the process called *multiplexing.* In a *time-division* multiplex system the signals corresponding to data from a particular sensor occupy a certain segment of the data sampling cycle. Time division can be achieved mechanically using a one-bar commutator that successively contacts the brushes connected to the various inputs. In an AM *frequency-division* multiplex system, each input modulates a subcarrier at a relatively low frequency. The modulated subcarriers are then combined and used to modulate the radio carrier. The signals corresponding to data from a particular sensor occupy a certain side band. At the receiver the carrier is demodulated and the signals separated by frequency-selective band-pass filters.

Tape Recorders

The magnetic tape recorder has completely changed the character of data processing. This widely used recording instrument has a great dynamic range over a wide frequency range with low distortion and great flexibility. The recorded signal can be stored permanently or played back immediately at the recording speed or a more convenient speed.

The operation of the recorder is indicated in Fig. 27.24. A flexible magnetic tape is moved past a *recording head* at constant speed by the *tape transport.* Variations in the signal current i produce variations in the flux density B at the gap and therefore variations in the magnetization of the tape material. To avoid

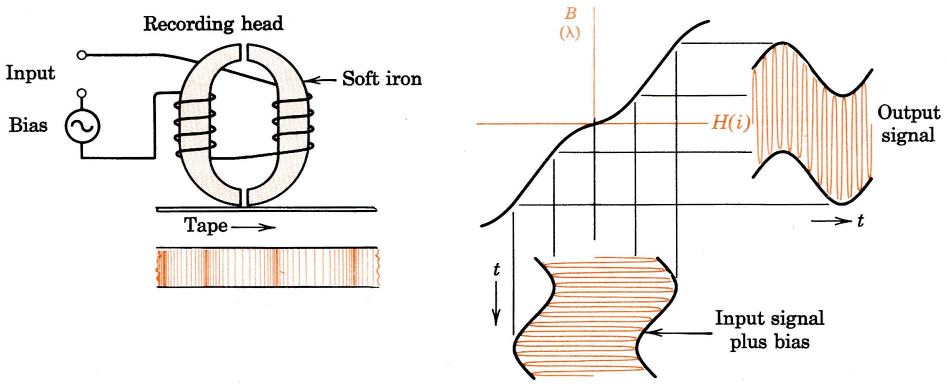


Figure 27.24 Magnetic tape recording.

distortion due to the nonlinearity of the magnetic circuit, the input signal is superimposed on a high-frequency constant-amplitude bias current. The peak values of the total signal (note that this is not amplitude modulation) lie on the linear portion of the magnetization curve. The *reproducing head,* identical with the recording head, produces a signal in response to the flux density variation. The high-frequency bias component is removed by a low-pass filter and the output signal resembles the input.

For a sinusoidal input current the peak value of the residual magnetism in the tape is of the form

$$B = B_0 + kI_m \cos \omega t + k_b I_b \cos \omega_b t \tag{27-16}$$

where I_b is the bias current. The emf induced in the reproducing head is

$$e = \frac{d\lambda}{dt} = NA\frac{dB}{dt} = k'\omega I_m \sin \omega t + k_b' I_b \omega_b \sin \omega_b t \tag{27-17}$$

After the bias frequency component is removed by filtering, the rms value of the signal is

$$E = k''fI \tag{27-18}$$

Since the voltage reproduced is directly proportional to frequency, the output amplifier must introduce a compensating frequency characteristic to provide an overall linear response.

Automatic Instrumentation Systems

Available now are completely automatic data handling systems that capture raw information in numerical form, evaluate it following a preset program using self-calibrated instruments, perform sophisticated data reduction, analyze the results, and print them out in the most usable form. They can handle up to 1000 channels bearing information on temperature, pressure, liquid level, strain, or any other physical parameter that can be converted to electrical form.

For speed and accuracy all data are processed in digital form, and visual readings and manual settings are eliminated. Sensors and instruments are programable and they are routinely calibrated by the same computer that performs the data reduction. Accuracies of 0.1% in absolute readings are possible, and data averaging (integration) procedures can further reduce errors and noise. Instead of the traditional strip chart, the results are displayed digitally or graphically, or they may be recorded on magnetic tape or punched cards for further analysis.

Instruments of the Future

As Barney Oliver points out,† the three revolutions in electronics—from vacuum tubes to transistors in the 1950s, from transistors to integrated circuits in the 1960s, and from ICs to large-scale integration in the 1970s—have been reflected in instrumentation. The availability of hundreds of active elements in a tiny chip at low cost permitted fresh design approaches that resulted in greatly expanded precision, sophistication, and reliability of instruments. Now the LSI microprocessor, a computer on a chip, promises a new area for exploitation. The capability and flexibility of large transducer–instrument–computer combinations can be incorporated in portable instruments along with autoranging, self-calibration, internal troubleshooting, and compensation for environmental changes. A possibility for the future is an open-ended instrument that accepts either preprogrammed or user-programmable ROM (read-only memory) instructions and into which the user enters his or her choice of parameters, functions, frequency ranges, data reduction operations, and display characteristics by means of a single keyboard as we now enter numbers in a hand-held calculator.

SUMMARY

◆ All branches of experimental engineering and science depend on instrumentation to extend human abilities to detect, observe, measure, control, compute, communicate, and display physical quantities.

◆ The d'Arsonval movement responds to direct current, but it can be used to measure voltage and resistance as well.
The effect of an instrument on the circuit being tested can be taken into account using the compensation theorem.
Ac quantities can be measured directly by iron-vane, thermocouple, or electrodynamometer instruments or indirectly by rectifier-type instruments.

---

† IEEE *Spectrum*, special issue on instrumentation, p. 44, November 1974.

- The electronic voltmeter employs electronic circuitry to provide high input impedance, accuracy, flexibility, and stability.
  The electronic counter employs a time base and gating circuit to determine the number of events per unit time or the time between events.
  The digital voltmeter uses analog-to-digital conversion to obtain better instrument characteristics and compatibility with digital equipment.
  The cathode-ray oscilloscope measures voltage and time, displays waveforms, compares phase and frequency, and identifies distortion.

- Instrumentation transducers convert physical variables to electrical form.
  Active transducers include: thermocouples, piezoelectric pickups, phototubes, photojunction cells, dynamic pickups, and tachometers.
  Passive transducers include: resistance thermometers, hot-wire anemometers, strain gauges, photoconductive cells, variable-reluctance elements, differential transformers, and liquid-level gauges.

- Bridge measurements based on null conditions are highly accurate.
  A slightly unbalanced bridge can be used as a transducer.
  Potentiometer measurements use voltage comparisons for high accuracy.
  Self-balancing potentiometers are valuable for recording and control.
  The strain gauge and bridge measures small displacements accurately.
  The differential transformer measures larger displacements accurately.

- In an instrumentation system all components must work together effectively. Such a system may include sensors and transducers, calibrating equipment, analog-to-digital and digital-to-analog converters, telemetering equipment, multiplexers, computers, recorders, and display devices.

REVIEW QUESTIONS

1. Define instrument sensitivity, error, accuracy, and precision.
2. How is a circuit affected by an ammeter? A voltmeter?
3. Why must an ammeter shunt selector switch "make" before "break"?
4. Explain the operation of an ohmmeter. A multimeter.
5. If an iron-vane instrument responds to i^2 and a rectifier instrument responds to I_{av}, how can they both read I_{rms}?
6. What are the advantages of an EVM over a dc voltmeter?
7. Explain the operation of an electronic counter.
8. Explain the operation of a DVM.
9. How can a CRO be used to measure voltage? Frequency? Phase?
10. Explain the operation of CRO sweep, trigger, and blanking circuits.
11. Distinguish between active and passive transducers; list 3 of each type.
12. Why are bridge and potentiometer measurements so accurate?
13. Explain the operation of an unbalanced bridge as a transducer.
14. Explain the operation of a self-balancing potentiometer.
15. How could a strain gauge be used to measure pressure?
16. How could a differential transformer be used to measure level?

EXERCISES

1. Design a multirange dc ammeter with ranges of 20 and 100 mA and 1, 2, and 5 A, using a 2-mA movement with 25-Ω internal resistance.
2. Design a multirange dc milliammeter with ranges of 1, 10, and 100 mA using a 100-μA movement with 10,000-Ω internal resistance.
3. Design a multirange voltmeter with ranges of 3, 10, 50, and 150 V using the movement of Exercise 1.
4. Repeat Exercise 3 with the movement of Exercise 2.
5. Design an ohmmeter, using the movement of Exercise 2 with a half-scale reading of: (a) 50 Ω (b) 5000 Ω.
6. Prove that the voltmeter reading in an ohmmeter changes most rapidly for a given percentage change in resistance R_X when $R_X = R_1$.
7. In the circuit of Fig. 27.25, an ammeter with an internal resistance of 200 Ω is:
   (a) inserted at point *a* and reads 50 mA, and then
   (b) inserted at point *b* and reads 50 mA.
   Estimate the original current in each case.

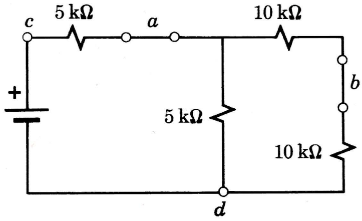


Figure 27.25

8. In the circuit of Fig. 27.25, a voltmeter with an internal resistance of 50 kΩ is
   (a) connected from point *a* to *c* and reads 4 V, and then
   (b) connected from point *d* to *b* and reads 4 V.
   Estimate the original voltage in each case.
9. Electrodynamometer, rectifier, and peak-reading instruments, all calibrated in rms values, are used to measure the current in Fig. 27.26. Predict the reading on each instrument.

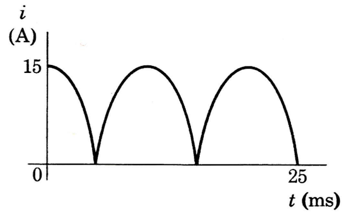


Figure 27.26

10. Repeat Exercise 9 for the voltage in Fig. 27.27.

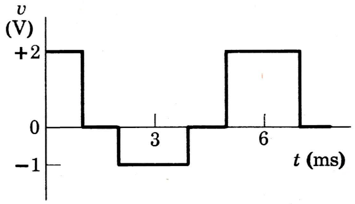


Figure 27.27

11. Repeat Exercise 9 for a voltage waveform (see Fig. 16.3b) consisting of a 10-V fundamental plus a 2-V third harmonic.
12. An amplifier has a bandwidth defined by lower and upper cutoff frequencies f_1 and f_2.
    (a) If a square wave of frequency f_1 is applied to the input, how are the fundamental and third harmonics affected in amplitude and phase? Draw the sum of these two components and compare with Fig. 27.9c.
    (b) Repeat part (a) for a square wave of frequency $f_2/3$ and compare with Fig. 27.9.
13. The speed of a rotating shaft is to be measured using a phototube and an electronic counter. Sketch a suitable arrangement.
14. In a DVM using the dual-slope technique, the reference voltage is 1.000 V and the

fixed "time" is 1000 counts. Predict the voltage indicated if the counter reads 762 on the down slope.

15. Devise an electrical circuit to measure the level of fuel in a tank using a parallel-plate capacitor.

16. In the Wheatstone bridge of Fig. 27.15, determine R_X when:

(a) $R_A = 1000\,\Omega$, $R_B = 400\,\Omega$, and $R = 144\,\Omega$.

(b) $R_A = 1000\,\Omega$, $R_B = 3000\,\Omega$, and $R = 729\,\Omega$.

17. Derive Eq. 27.13.

18. In the Maxwell bridge of Fig. 27.17, $R_A = 1000\,\Omega$, and $C_B = 0.025\,\mu\text{F}$. $R_B = 1500\,\Omega$ and $R = 240\,\Omega$ for a balance at 1000 Hz. Determine the parameters of the unknown.

19. In a slidewire potentiometer (Fig. 27.19), a standard cell provides $V_S = 1.020$ V and $R_S = 2040\,\Omega$.

(a) To measure potentials up to approximately 10 V, specify values of R_P, R_A, and V.

(b) For $R_O = 1520\,\Omega$, determine the unknown voltage.

20. In Fig. 27.18a, $R + \Delta R$ is a strain gauge ($K_G = 3$). The dc voltage source is replaced by a voltage $v_S = 20 \cos 10^5 t$ V and $R = 200\,\Omega$. The output voltage is observed to be $v = 6 \sin 20\pi t \cos 10^5 t$ mV. Determine the amplitude and frequency of the strain sensed by the gauge.

21. A differential transformer has the characteristics of Fig. 27.22b where the voltage scale is 1 V/cm and the displacement scale is 0.02 in./cm. (The excitation of the primary is 10 V rms at 5 kHz.) If the displacement of the core is $0.01 + 0.002 \cos 20\pi t$ in., predict the output voltage and percentage modulation.

Appendix

Physical Constants
Conversion Factors
Device Characteristics
Color Code for Resistors
Preferred Resistance Values
Answers to Selected Exercises

Physical Constants†

Charge of the electron	e	1.602×10^{-19} C
Mass of the electron (rest)	m	9.109×10^{-31} kg
Velocity of light	c	2.998×10^{8} m/s
Acceleration of gravity	g	9.807 m/s^2
Avogadro's number	N_A	6.023×10^{23} atoms/g-atom
Planck's constant	h	6.626×10^{-34} J-s
Boltzmann's constant	k	8.62×10^{-5} eV/K
Permeability of free space	μ_o	$4\pi \times 10^{-7}$ H/m
Permittivity of free space	ϵ_o	$1/36\pi \times 10^{9}$ F/m

Conversion Factors

1 inch = 2.54 cm
1 meter = 39.37 in.
1 angstrom unit = 10^{-10} m
1 kilogram = 2.205 lb (mass)
1 newton = 0.2248 lb (force)

1 hertz = 1 cycle/s
1 electron volt = 1.60×10^{-19} J
1 joule = 10^{7} ergs
1 tesla = 1 Wb/m^2
= 10^{4} gauss

---

† E. A. Mechtly, "The International System of Units: Physical Constants and Conversion Factors," NASA SP-7012, Washington, D.C., 1964.

Device Characteristics

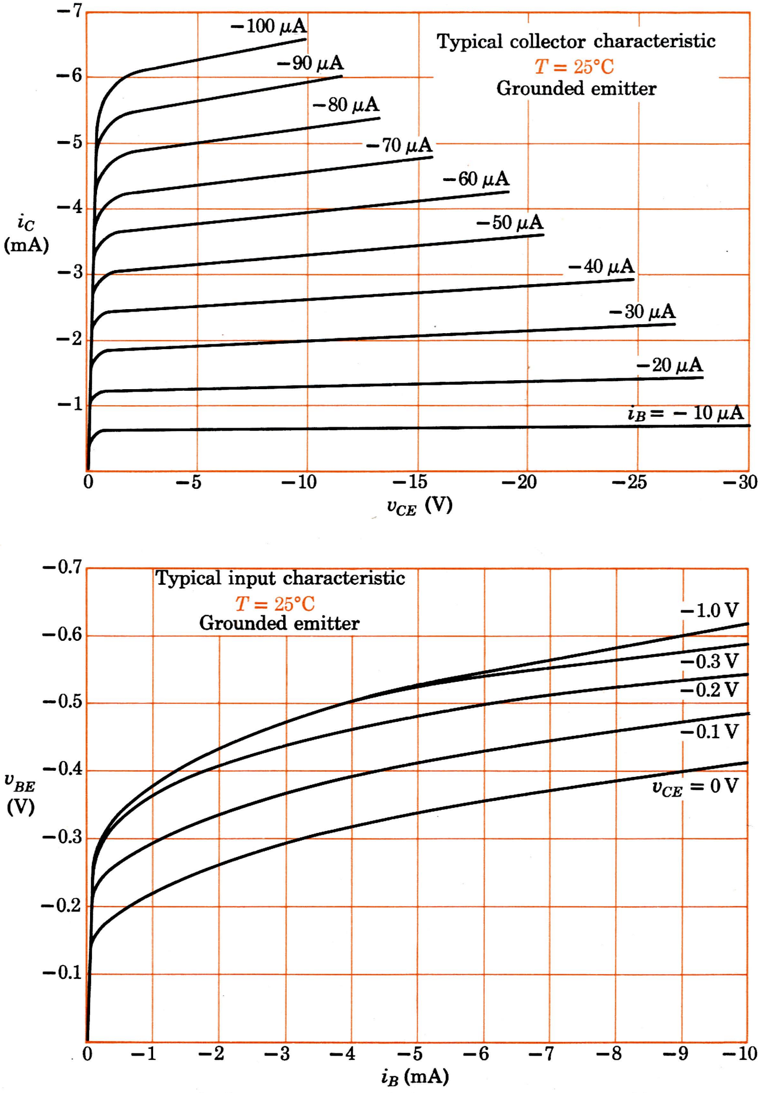
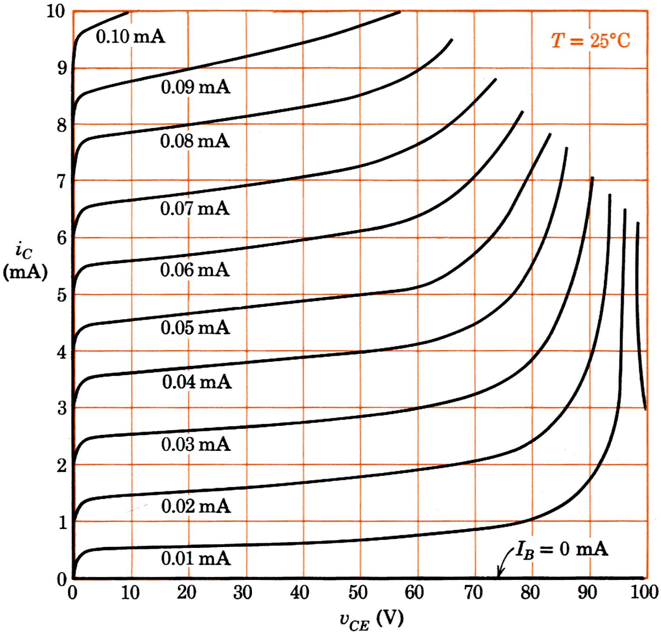
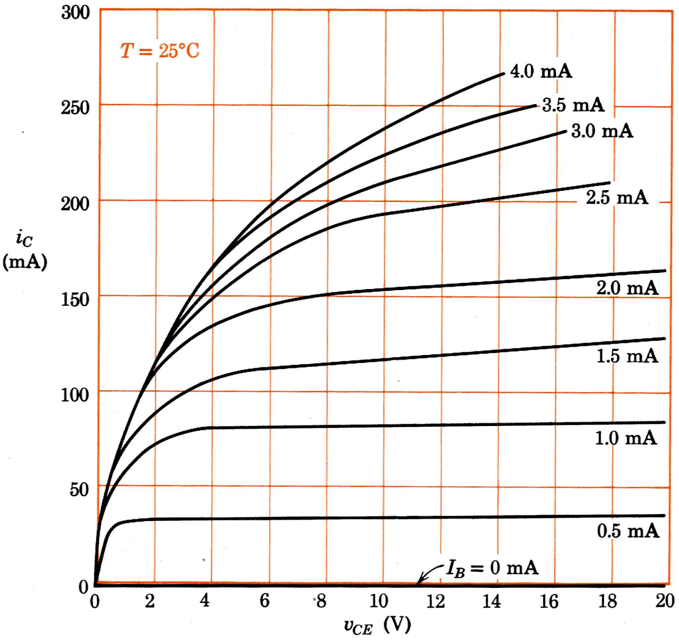


Figure A1 Typical characteristics of the 2N502 transistor. (Courtesy Philco Corp.)



(a) High voltage region



(a) Saturation region

Figure A2 Collector characteristics for 2N699B transistor. (Courtesy Fairchild Semiconductor)

2N3114 *NPN* High-Voltage, Medium-Power Silicon Planar† Transistor

MAXIMUM RATINGS

Operating junction temperature	200°C
Power dissipation: 25°C case	5 W
25°C free air	0.8 W
V_{CBO} Collector-to-Base voltage	150 V
V_{EBO} Emitter-to-Base voltage	5 V

TYPICAL SMALL-SIGNAL CHARACTERISTICS

($f = 1$ kHz, $I_C = 1$ mA, $V_{CE} = 5$ V)

h_{fe}	Small-signal current gain	50
h_{ie}	Input resistance	1,500 Ω
h_{oe}	Output conductance	5.3 μ℧
h_{re}	Voltage feedback ratio	1.5×10^{-4}
h_{ib}	Input resistance	27 Ω
h_{ob}	Output conductance	0.09 μ℧
h_{rb}	Voltage feedback ratio	0.25×10^{-4}

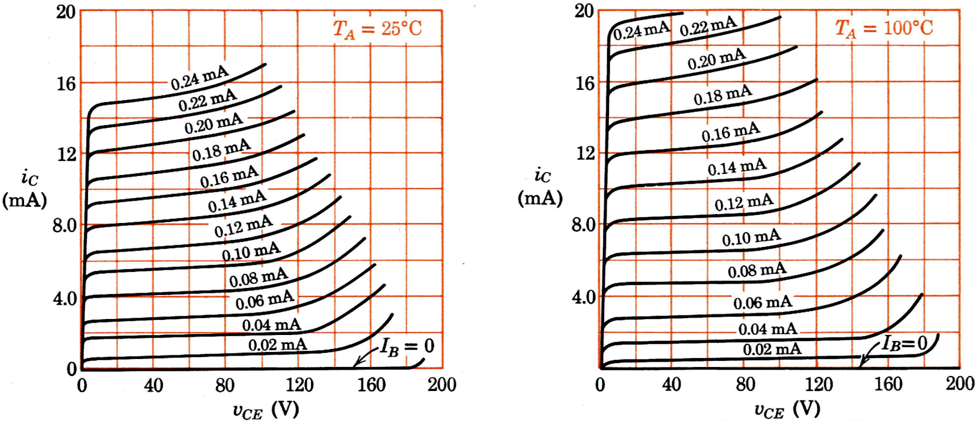
ELECTRICAL CHARACTERISTICS (25°C free air)

Symbol	Characteristic	Min.	Typical	Max.	Test conditions
V_{BE}(sat)	Base saturation voltage		0.8	0.9 V	$I_C = 50$ mA $I_B = 5.0$ mA
V_{CE}(sat)	Collector saturation voltage		0.3	1 V	$I_C = 50$ mA $I_B = 5.0$ mA
I_{CBO}	Collector cutoff current		0.3	10 nA	$I_E = 0$ $V_{CB} = 100$ V
I_{CBO}(150°C)	Collector cutoff current		2.7	10 μA	$I_E = 0$ $V_{CB} = 100$ V
$h_{FE}(\beta_{dc})$	DC current gain	15	35		$I_C = 100$ μA $V_{CE} = 10$ V
h_{fe}	20-MHz current gain	2.0	2.7		$I_C = 30$ mA $V_{CE} = 10$ V
C_{ob}	Output capacitance		6	9 pF	$I_E = 0$ $V_{CB} = 20$ V
C_{TE}	Emitter transition capacitance		70	80 pF	$I_C = 0$ $V_{EB} = 0.5$ V

Figure A3 Typical characteristics of the 2N3114 transistor. (Courtesy Fairchild Semiconductor)

---

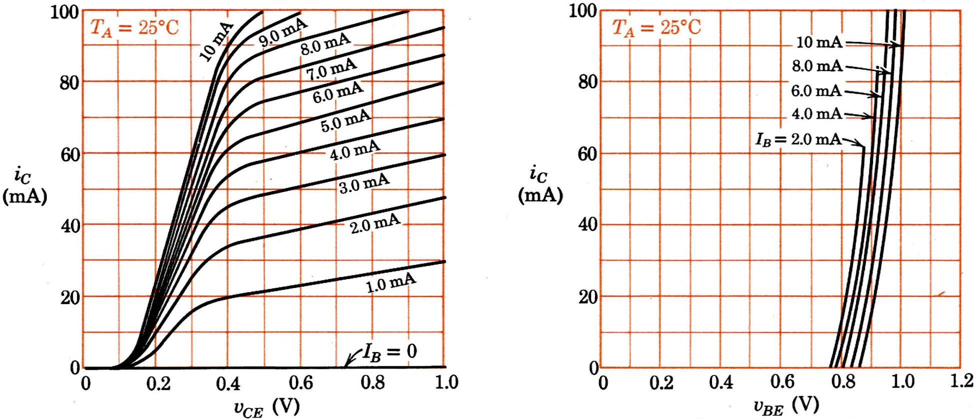
†Planar is a patented Fairchild process.



$T_{\text{Ambient}} = 25°\text{C}$

$T_{\text{Ambient}} = 100°\text{C}$

(a) High voltage region



Collector characteristics

Base characteristics

(b) Saturation region

Figure A3 (continued)

μA741 Frequency Compensated Operational Amplifier

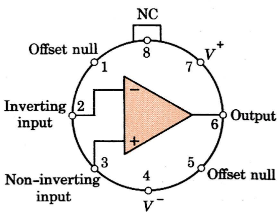
General Description

The μA741 is a high performance monolithic operational amplifier constructed on a single silicon chip, using the Fairchild Planar† epitaxial process. It is intended for a wide range of analog applications. High common mode voltage range and absence of "latch-up" tendencies make the μA741 ideal for use as a voltage follower. The high gain and wide range of operating voltage provides superior performance in integrator, summing amplifier, and general feedback applications. The μA741 is short-circuit protected, has the same pin configuration as the popular μA709 operational amplifier, but requires no external components for frequency compensation. The internal 6dB/octave roll-off insures stability in closed loop applications.

Absolute Maximum Ratings

Supply voltage	±22 V
Internal power dissipation (note 1)	500 mW
Differential input voltage	±30 V
Input voltage (note 2)	±15 V
Voltage between offset null and V^-	±0.5 V
Storage temperature range	−65°C to +150°C
Operating temperature range	−55°C to +125°C
Lead temperature (soldering, 60 s)	300°C
Output short-circuit duration (note 3)	Indefinite

Connection diagram
(Top view)


NC
Offset null
8
V^+
1
7
Inverting input
2
−
+
Output
6
3
5
Non-inverting input
4
Offset null
V^-

Note: Pin 4 connected to case

Notes

1. Rating applies for case temperatures to 125°C; derate linearly at 6.5 mW/°C for ambient temperatures above +75°C.
2. For supply voltages less than ±15 V, the absolute maximum input voltage is equal to the supply voltage.
3. Short circuit may be to ground or either supply. Rating applies to +125°C case temperature or +75°C ambient temperature.

Equivalent circuit

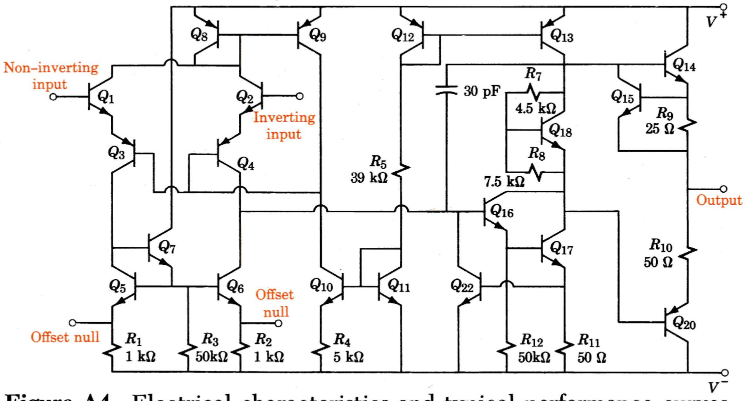
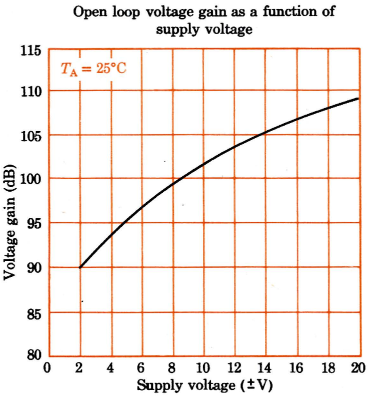

V^+
Q_8
Q_9
Q_{12}
Q_{13}
Q_{14}
Non-inverting input
Q_1
Q_2
Inverting input
30 pF
R_7
4.5 kΩ
Q_{15}
R_9
25 Ω
Q_{18}
Q_3
Q_4
R_5
39 kΩ
R_8
7.5 kΩ
Output
Q_{16}
Q_7
Q_{17}
R_{10}
50 Ω
Q_5
Q_6
Offset null
Q_{10}
Q_{11}
Q_{22}
Q_{20}
Offset null
R_1
1 kΩ
R_3
50kΩ
R_2
1 kΩ
R_4
5 kΩ
R_{12}
50kΩ
R_{11}
50 Ω
V^-

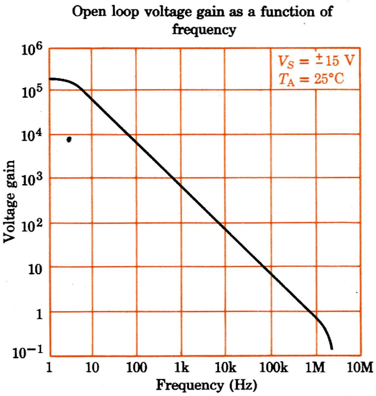
Figure A4 Electrical characteristics and typical performance curves for the μA741. (Courtesy Fairchild Semiconductor.)

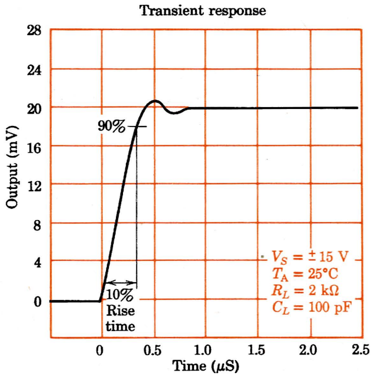
Electrical Characteristics ($V_S = \pm 15$ V, $T_A = 25°$C unless otherwise specified)

Parameters (see definitions)	Conditions	Min.	Typ.	Max.	Units
Input offset voltage	$R_s \leq 10\ k\Omega$		1.0	5.0	mV
Input offset current			20	200	nA
Input bias current			80	500	nA
Input resistance		0.3	2.0		MΩ
Input capacitance			1.4		pF
Offset voltage adjustment range			±15		mV
Large-signal voltage gain	$R_L \geq 2\ k\Omega$ $V_{out} = \pm 10$ V	50,000	200,000		
Output resistance			75		Ω
Output short-circuit current			25		mA
Supply current			1.7	2.8	mA
Power consumption			50	85	mW
Transient response (unity gain)	$V_{in} = 20$ mV $R_L = 2\ k\Omega$ $C_L \leq 100$ pF				
Rise time			0.3		µs
Overshoot			5.0		%
Slew rate	$R_L \geq 2\ k\Omega$		0.5		V/µs
CMRR	$R_s \leq 10\ k\Omega$	70	90.0		dB

Typical Performance Curves


Open loop voltage gain as a function of supply voltage


Open loop voltage gain as a function of frequency


Transient response

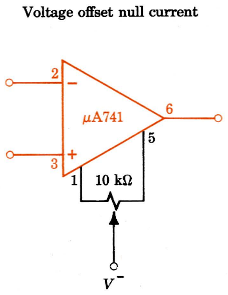

Voltage offset null current

Figure A4 (continued)

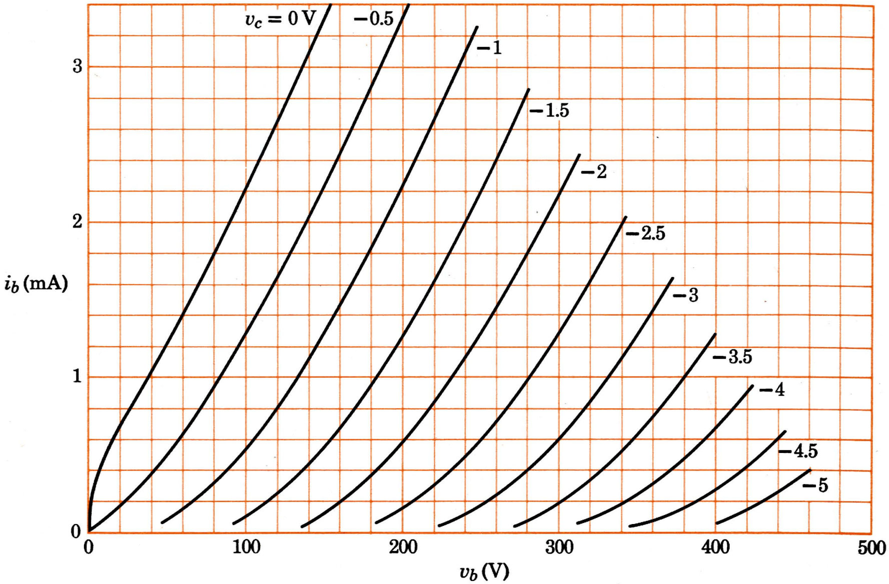


Figure A5 Average plate characteristics of the 12AX7 triode. (Courtesy General Electric Co.)

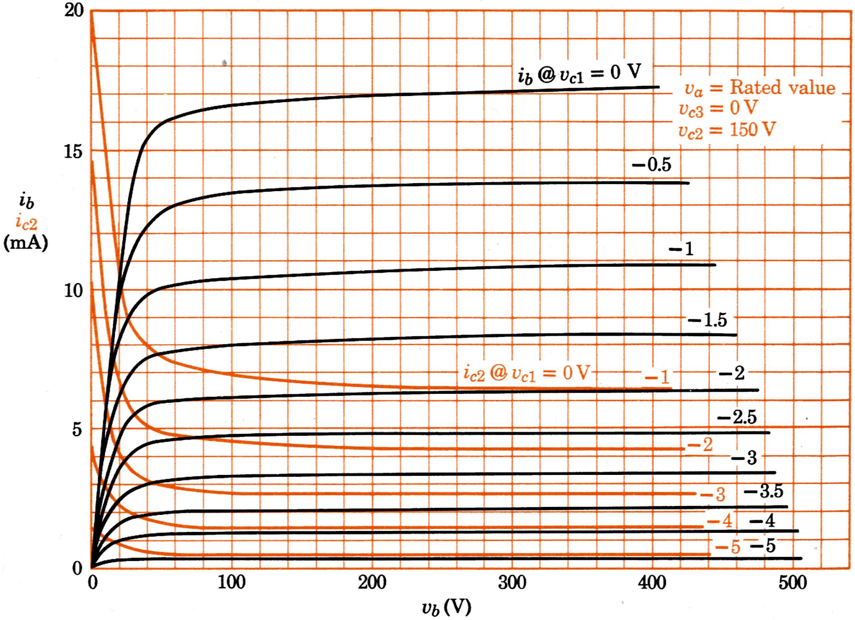


Figure A6 Average plate characteristics of the 6AU6-A pentode. (Courtesy General Electric Co.)

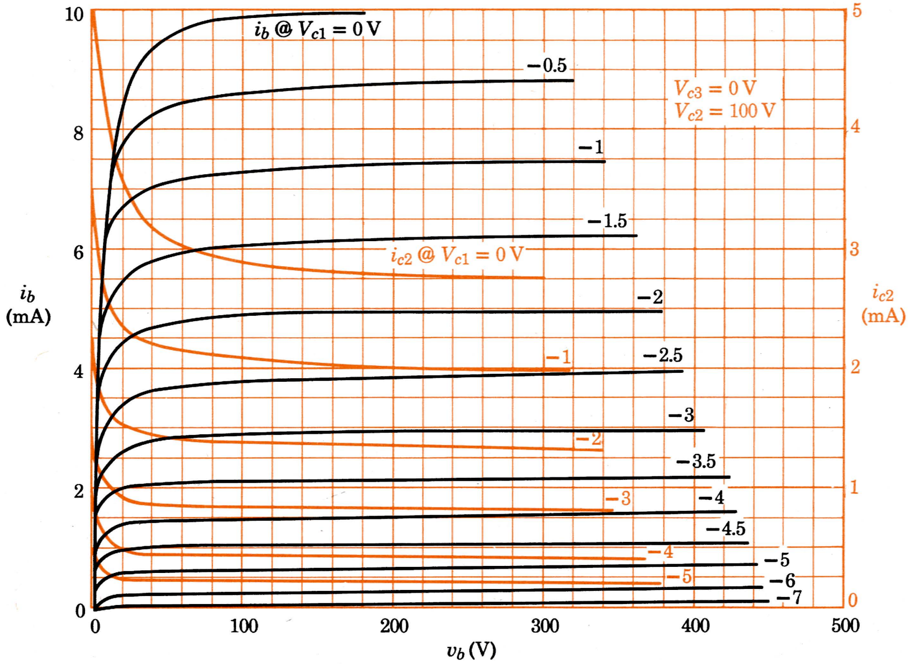


Figure A7 Average plate characteristics of the 6SJ7 pentode. (Courtesy General Electric Co.)

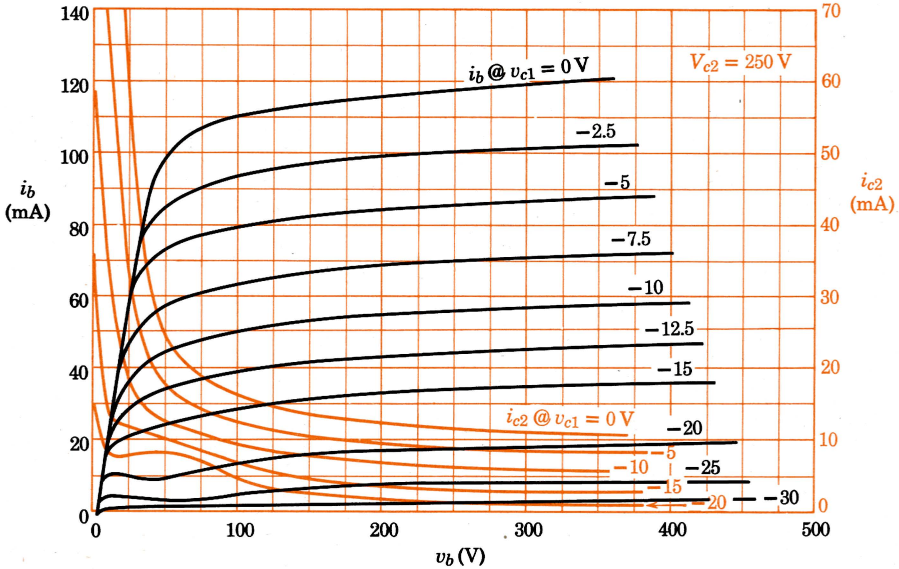
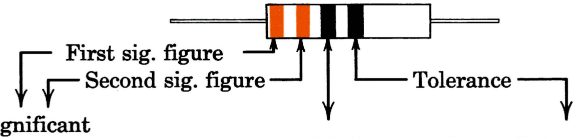


Figure A8 Average plate characteristics of the 6AQ5-A beam-power pentode. (Courtesy General Electric Co.)

Color Code for Resistors

Ordinary composition fixed resistors are characterized by power dissipation in watts, resistance in ohms, and tolerance in percent. They vary in size from 0.067 in. (0.1 W) to 0.318 in. (2 W) in diameter. For clarity, the resistance and tolerance values are indicated by color bands according to a standard code. (For example, bands of orange, orange, black, and black would identify a 33 × 1 or 33-Ω, 20%-tolerance resistor.)



Significant figures	Color	Multiplier	Color	Tolerance
0	Black	1	Black or no color	±20%
1	Brown	10		
2	Red	100	Silver	±10%
3	Orange	1,000	Gold	±5%
4	Yellow	10,000		
5	Green	100,000		
6	Blue	1,000,000		
7	Violet			
8	Gray			
9	White			
	Gold	0.1		
	Silver	0.01		

Example shown:
33 Ω ±20%

A similar color scheme is used to indicate the value and tolerance of small capacitors; an extra color band indicates the temperature coefficient.

Preferred Resistance Values

To provide the necessary variation in resistance while holding the number of resistors to be stocked to a reasonable minimum, the electronics industry has standardized on certain preferred values. The range in significant figures from 1.0 to 10 has been divided into 24 steps, each differing from the next by approximately 10% ($\sqrt[24]{10} = 1.10$). The preferred values, available in all multiples in ±5% tolerance resistors, are:

1.0‡	1.5‡	2.2‡	3.3‡	4.7‡	6.8‡
1.1	1.6	2.4	3.6	5.1	7.5
1.2†	1.8†	2.7†	3.9†	5.6†	8.2†
1.3	2.0	3.0	4.3	6.2	9.1

† Also available in ±10% tolerance resistors.

‡ Also available in ±10% and ±20% tolerance resistors.

ANSWERS TO SELECTED EXERCISES

CHAPTER 1

1. (a) 102 kg; (b) 4572 J, 0.00127 kWh;
   (c) 1524 W.
3. 50 kV, 0.25 J.
7. (a) 4.0×10^{-15} J; (b) 0.08 mA.
10. $v_R = 2$ V, $v_L = 15$ V.
14. (a) 0.5 mJ; (b) 5 mJ.
17. (a) 1500 W; (b) 500 W; (c) 2000 W;
    (d) 1000 J; (e) 500 J.
19. (a) $v_C = V$, $i_L = 0$; (b) $i_2 = V/R_2$

CHAPTER 2

2. 2 A, 8 V, 4 V.
5. 3 A, 2 A, −3 A.
8. $v_{cb} = +10$ V.
9. (a) $5 + 20 \cos t$ A, $-80 \sin t$ V.
12. 2 A.
15. 4 A.
17. 3.93 V.
19. 1 A.
22. $v_{ab} = -6$ V.
25. (b) $9.1 + 0.3 \cos \omega t$ mA.
28. 10.2 V, 1.6 mA.

CHAPTER 3

3. $i(t) = 16.9\, e^{-t/3}$ mA, 0.433 mA.
6. $i_R = 4\, e^{-200t}$ mA, $i_C = -120\, e^{-200t}$ mA,
   $i_L = -30\, e^{-200t}$ mA, $i = -146\, e^{-200t}$ mA.
9. $i(t) = 10 \cos (10^3 \pi t/6 + 2\pi/3)$ A, 0 A,
   $T = 12$ ms, $f = 83.3$ Hz.
12. (a) $2\sqrt{2}\,\underline{/-15^\circ}$; (b) $1 - \mathrm{j}1$; (c) $1\underline{/-150^\circ}$;
    (d) $2\sqrt{10}\,\underline{/11.6^\circ}$; (e) $4\underline{/150^\circ}$;
    (f) $4.46 + \mathrm{j}3$; (g) $2\underline{/90^\circ}$; (h) $\pm 2\underline{/15^\circ}$;
    (i) $\sqrt[6]{2}\,\underline{/15^\circ}$, $\sqrt[6]{2}\,\underline{/135^\circ}$, $\sqrt[6]{2}\,\underline{/255^\circ}$.
15. (a) $10\underline{/-126^\circ}$, $-5.88 - j8.1$;
    (b) $5\underline{/-90^\circ}$, $-\mathrm{j}5$; (c) $2\underline{/-45^\circ}$,
    $\sqrt{2} - \mathrm{j}\sqrt{2}$; (d) $50\underline{/0^\circ}$, $50 + \mathrm{j}0$.
18. (a) $\mathbf{I} = 3.22\underline{/97.1^\circ}$,
    $\mathrm{i(t)} = 3.22 \cos (\omega \mathrm{t} + 97.1^\circ)$.
21. $V_{\mathrm{av}} = VD$, $V_{\mathrm{eff}} = VD^{1/2}$.
24. 29.8 → 30-hp motor.

CHAPTER 4

3. (a) $i_{C1} = 0$, $i_{R1} = 0$, $v_{C1} = V$;
   (c) $i(t) = (V/R)\, e^{-2t/RC}$.
6. (a) $v_L = 0$, $i_{R2} = V/R_2$; (b) $i = V/R_2$;
   (c) $i(t) = (V/R_2)\, e^{-(R_1+R_2)t/L}$;
   (d) $T = L/(R_1 + R_2)$.
9. (a) 12.5 ms; (b) $v_{C1} = 15\, e^{-80t}$ V,
   $i_{R2} = 3\, e^{-80t}$ mA.
12. (a) $i = 0$, $v_R = 0$, $v_L = 25$ V;
    (b) $s^2 L + sR + 1/C = 0$;
    (e) $i(t) = 33.3(e^{-4t} - e^{-16t})$ mA.
14. (e) $i(t) = 3\, e^{-60t} - e^{-20t}$ A.
17. (b) $i(t) = 2\, e^{-2t} - 4t\, e^{-2t}$ A.
18. (e) $v(t) = 10.54\, e^{-t} \cos (3t - 18.4^\circ)$ V.

23. (a) $Z(s) = (8 + 4s)/(6 + 2s)$; (b) $4/3\,\Omega$; (c) $0.25\,e^{-4t}$, 0, ∞; (d) $8\,e^{-4t}$ V.

26. (a) $\dfrac{R_1R + sL(R + R_1)}{R + sL}$, $\dfrac{(sCR + 1)R_1}{sC(R_1 + R) + 1}$.

29. (c) $v(t) = A\,e^{-t}\cos(3t + \varphi)$.

31. (b) $i(t) = 4\,e^{-t/2} - e^{-2t}$ A.

CHAPTER 5

1. $2\,e^{-2t}$ A, $-8\,e^{-2t}$ V, $2.8\,e^{-2t}$ A.
4. 2 A.
6. (a) 60 V; (b) $-10\,e^{-4000t}$; (c) $23.6\cos(4000t - 21.9^\circ)$ V.
9. (a) $V_m\sqrt{(1/R)^2 + \omega^2C^2}\cos(\omega t + \tan^{-1}\omega CR)$.
12. (a) $i(t) = 14.14\sqrt{2}\cos(1000t - 5^\circ)$ mA.
14. (a) $R_P = 1133\,\Omega$, $C_P = 1.48\,\mu$F; (b) $R_P = 508\,\Omega$, $C_P = 0.82\,\mu$F.
18. (b) $10\underline{/0^\circ}$ A, $28.3\underline{/45^\circ}$ V, $14.1\underline{/135^\circ}$ A, $10\underline{/90^\circ}$ A; (c) $10\cos(1000t + 90^\circ)$ A.
22. $50\sqrt{3}\,\Omega$.
25. $20\underline{/0^\circ}$, $15\underline{/90^\circ}$ A.
28. (a) $x(t) = X_o - 0.5\cos 5t$ m; (b) $x(t) = X_o - 0.4\cos(5t + 37^\circ)$ m.
30. (b) $i(t) = G_1v_a + C\,d(v_a - v_b)/dt$, $0 = G_2v_b + (1/L)\int v_b\,dt + C\,d(v_a - v_b)/dt$.

CHAPTER 6

2. (a) $L\,di/dt + Ri = V$.
5. $i(t) = (V + 5)(1/R)\,e^{-t/RC} = 12.5\,e^{-100t}$ mA, $V_C(t) = V - (V + 5)\,e^{-t/RC} = 20 - 25\,e^{-100t}$ V.
9. 50 V, 3.51 kΩ.
11. (a) $2\cos(1000t - 60^\circ) - e^{-577t}$ A; (b) $t = 2.62$ ms.
14. $1.72\cos(5t - 16^\circ) - 0.64\,e^{-6.7t}$ A.
17. (b) $0.2\,(e^{-100t} - e^{-200t})$ A.
20. $15\,t\,e^{-5t}$ A.
23. $-66.5\,e^{-66t} - 0.66\,e^{-300{,}000t} + 66.7\,(\cos 1000t + 4^\circ)$ mA.
27. For a pulse, $I_o = M_P/L = 5$ mA; $i(t) \cong 5\cos 1000t$ mA.

CHAPTER 7

3. (a) 10 A, 300 V; (c) 50 Ω.
5. (a) 1100 W; (b) 35.4 W.
9. 113 kVA, 0.707.
11. (a) $64 + j28 = 69.9\underline{/23.6^\circ}$ kVA; (c) 35 A; (d) 8 kVAR capacitive.
14. $\omega_{co}L = R\|R_L$; for $R = 10\,\text{k}\Omega$, $L \cong 0.8$ H.
18. 540 – 1600 kHz, about 100 – 900 pF.
21. 12.7 kHz, 100 mA, 100 Ω, 6.25 nF, 200 V.
24. $L = 49.8\,\mu$H, $C = 127$ pF, $G = 20\,\mu\mho$.
27. (b) 10 krad/s, 10, 1 krad/s; (c) 1 Ω, 1 mH, 10 μF.
30. (a) 3.12 kHz; (c) 500 Ω.
34. (a) $6\underline{/120^\circ}$, $6\underline{/0^\circ}$, $6\underline{/-120^\circ}$ A; (b) 0; (c) $416\underline{/60^\circ}$ V.
37. (b) $I_P = 16$ A, $I_1 = 27.7$ A; (d) 9600 W.
40. 120 V.
43. 1.61 Ω, 10 kW.
45. 55.45 A.

CHAPTER 8

2. For $R_{eq} = R_1 + R/2$, $T = L/R_{eq}$, $V_T = V/2$; $i_L(t) = (V_T/R_{eq})(1 - e^{-R_{eq}t/L})$.
6. 2500 W.
8. 0.6, -0.4, -0.4, 0.6 $\mho$; 0.2, 0.4, 0.2 $\mho$.
11. $Z_{31} = \infty$, $Z_{12} = 1\,\Omega$, $Z_{23} = 2\,\Omega$.
14. $0.995\cos(10t + 84^\circ) + 9.5\cos(1000t + 6^\circ)$; Hi-pass filter.
17. 1 mH, 5 Ω, 12.5 W.
20. (a) 50 Ω; (b) 10 Ω; (c) 0.
22. (a) 5 mH; (b) 8 and 32 mH.
24. $1250\underline{/-90^\circ}$ V, $417\underline{/90^\circ}$ Ω.

CHAPTER 9

2. (b) $BC/[1 - B(C - A)]$.
5. $G = AB$, $H = C - 1/B$, $G_F = AB/(1 + A - ABC)$.
8. $\mathbf{C} = -0.001 - j\,0.05$.
12. (a) $T(s) = X_2(s)/X_1(s) = (1/KD)/(s + 1/KD)$; (b) Pole at $s = -1/KD$, Zero at ∞; (c) $x_2(t) = X_1(1 - e^{-t/KD})$.
15. $T(j\omega) = 10/(25 - \omega^2)$.
17. (a) $10/(s + 2)$; (b) $6s/(s + 0.02)$.
20. (a) $T(j\omega) = 1/(0.1 + j\omega)$; (b) $y(t) = 10 - 10\,e^{-t/10}$.

CHAPTER 10

2. (a) 3640 V/m, 1250 ns; (b) 335 kV, 1250 ns; (c) 2.9×10^{-17}, 5360×10^{-17} J.
5. $x = 3.65$ cm, $y = 5$ cm.
8. $m = r^2B^2e/2V$.
11. 4.76×10^{-10} s, 33 MHz.
14. Vert 0.2 V/cm, Horiz 1 ms/cm.
17. Thorium, by $4.83 \times 10^7 : 1$.
19. 8.4×10^5 m/s.

22. (a) 70 MHz; (b) longer.
25. 19.6 mA.
28. $A_V \cong 90$.

CHAPTER 11

2. 22.2 A.
4. (a) Joules, 39.6 V^{-1};
   (b) rate decreases exponentially with energy gap.
7. 2×10^9 atoms/e-h pair, $\rho \cong 0.45\ \Omega$-m.
10. (a) 230 V; (b) 92 kV.
12. (a) 3.33×10^{12}; (b) p-type;
    (c) $5 \times 10^{20} \gg 4 \times 10^{11}$, yes; (d) 22 mV;
    (e) 10,400.
16. $1.56 \times 10^{23}\ m^3$.
18. 38,600 A/m^2.
22. (a) -0.098 V; (b) 0.98 mA.
25. By graphical solution, $I \cong 16$ mA.

CHAPTER 12

2. (a) $R_F \cong 2.8\ k\Omega$; (b) 2.7 mA.
5. Forward: ideal diode;
   reverse: ideal diode, $V_Z = 19.6$ V,
   $R_Z = 2\ \Omega$.
7. (a) Reversed ideal diode, $V = -2$ V,
   $R = 17\ \Omega$ in series; (b) Ideal diode in parallel with $I_s = 60$ mA and
   $G = 20\ m\mho$.
10. 54 mA, 108 V, 7.2 W.
12. (a) 33.8%; (b) 58%; (c) 67.7%
16. 215 V, 0.008, 13.5°, 1.07 A.
19. 1%, 0.5%.
21. (a) 222 V; (b) 8.27 H; (c) 0.0213;
    (d) 0.06 V;
24. (a) 51 V, 31 V; (b) 1.4%; (c) 30.8 V, 0.6%;
    (d) 0 A, 22 V.
27. v_2 clipped at $+2$ and -5 V.
30. (a) For $v_1 < 4$, $v_2 = 4$; for $v_1 > 4$,
    $v_2 = v_1$; (b) For $v_1 < 4$, $v_2 = v_1$; for
    $v_1 > 4$, $v_2 = 4$.
33. v_2 clamped at $V_B - v_C = 0$ V.

CHAPTER 13

2. $i_d \cong 2 \sin \omega t$ mA, $A_V \cong 10$.
5. (b) 4 kΩ; (c) -6 V in ser. with 900 Ω.
10. (b) -0.1 mA; (c) -6 mA, $+6.1$ mA,
    -6 V.
14. (a) 1000 yr; (b) 10 yr.
17. $A = 37.5\ mil^2$.
18. $f \cong 670$ Hz.
21. (a) 4.77 A, 746 W; (b) 2.36 A, 336 W;
    (c) 4.8 W.

CHAPTER 14

3. (a)

A	B	C	f
0	0	0	0
0	0	4	1
0	4	0	1
0	4	4	2
4	0	0	1
4	0	4	2
4	4	0	2
4	4	4	3

OR, MAJORITY, AND

(b) 2.5 V.

6. 2^n.
8. 4.98 V, 0.3 V.
12. $R_{ON} \cong 6\ \Omega$, $R_{OFF} \cong 460\ M\Omega$; 21 Ω, 15 kΩ.
14. (a) 3.75 V; (b) 0.3 V.
17. $V_{B3} \cong 0.7$, $V_{B2} \cong 1.4$, $V_{B1} \cong 2.1$;
    $V_{B4} \cong 1.0 < 1.7 \therefore$ OFF.
22. (a) 0.3, 5; (b) 0.3, 5; (c) 5, 0.3; (d) 5, 0.3;
    (e) 0.3, 5 V.
25. Output $\cong 10$ V and $\cong 0$ V; INVERSION.
27. A 500-Hz square wave.

CHAPTER 15

1. (a) 00011, 01010, 10100, 11111; (b) 1,
   4, 10, 26.
3. 16.25, 1.5; 10000.01, 1.1; 24.375, 10.8$\overline{3}$;
   11000.011, 1010.1$\overline{10}$.
6. 00011.
10. $A \cdot B = \overline{\overline{A} + \overline{B}} = \overline{\overline{A + 0} + \overline{B + 0}}$.
12. $f = \overline{\overline{A}CD}$.
14.

A	B	AB	$\overline{AB}$	OR	NOR	XOR	EQ
0	0	0	1	0	1	0	1
0	1	0	0	1	0	1	0
1	0	0	0	1	0	1	0
1	1	1	0	1	0	0	1

17. Use NAND, OR, NOR, and AND gates.
20. (a) $f = A\overline{B} + \overline{A}B$; (b) $\overline{\overline{A\overline{B}} \cdot \overline{\overline{A}B}}$.
22. $\overline{B}\overline{D}$.
25. (c) f: 5 gates, 13 leads, 2 levels,
    $\overline{f}$: 4 gates, 10 leads, 3 levels.

CHAPTER 16

2. (a) 2 mA, 8 V; (c) $0.55 \sin \omega t$ mA; (d) 6.6.
5. (a) 2 mA, 8 V; (b) 1.8 mA, 9.6 V.
8. (a) -6.5 V; (b) 4.75 kΩ.
11. (a) 13 V; (b) 1.95 V, 10.2 kΩ; (c) 1.63 mA.
14. (a) $R_E = R_C = 6\ k\Omega$, $R_1 = 18\ k\Omega$,
    $R_2 = 33\ k\Omega$; (b) $0.46 < I_C < 0.51$ mA.
17. (a) $v_{CE}(\text{sat}) = 0$, $I_{CEO} = 0$, $V_{CE} = 25$ V,
    $I_C = 400$ mA; (b) $P_o = 5$ W, $\eta = 25\%$.

19. $N_2/N_1 \cong 3.5$, $\eta \cong 50\%$.
22. (b) One possibility: $V_{CE} = 25$ V, $I_C = 250$ mA; (c) 3.125 W, 3.125 W, 25%; (d) $N_2/N_1 \cong 1.25$ 10 W, 5 W, 50%.

CHAPTER 17

3. 2000, 20 krad/s.
6. (a) 0, 3, 200, 203, 206 krad/s; (c) If load responds to audio only.
7. $r_{ac} = 1/(1 + 0.02\,v) = 0.83\,\Omega$. $i(t) = 11 + 1.2 \sin \omega t$.
12. (a) 3000 $\mu\mho$; (b) 2000 $\mu\mho$.
16. 4.5 mA, 2.5 mV.
19. (a) npn; (b) yes; (d) $\alpha R_L/r_e$; (e) 39.6.
22. (a) 5 mV; (b) 9 mV.
25. $720 < r_\pi < 3030\,\Omega$; no.

CHAPTER 18

2. (a) 10 kΩ; (b) $-30\angle\underline{-53^\circ}$.
5. (a) $R_D \| R_L \| r_d = R_D$, $R_s + R_1 \| R_2 = R_1 \| R_2$; (b) -10.
7. (a) 0.65 mA, 4 kΩ; (b) 6 kΩ; (c) 115; (d) 72.
9. (b) 4.4 kΩ; (c) 1.5 kΩ, 17 V.
12. (a) 16 Hz, 81 Hz; (b) 81 Hz; (c) 1 μF for $R_L = r_\pi \| R_B \cong 1$ kΩ.
15. (b) $C_{be} = 100$ pF; (c) 6.4 MHz.
18. 64 kHz.
21. -50, 80 kHz.
24. (a) 62 dB; (b) 72 dB, 52 dB.
28. $f_2 = 0.6$ MHz, 0 dB, 23.5 dB.
30. 3.
33. $\mathbf{G} = 200\angle\underline{0^\circ}$.
36. $0.0067\angle\underline{180^\circ}$.
40. $H = -0.043$.
42. (a) $H = -0.095$.
44. 1.9 V, $R_i = 73$ kΩ, $R_o = 53\,\Omega$.
46. (a) 0.32 MHz; (b) 3.2 MHz.

CHAPTER 19

2. (a) $R_1 = 5$ kΩ, $R_F = 1$ MΩ, inverting; (c) for $R_1 = 5$ kΩ, $R_F = 95$ kΩ, noninverting.
5. (a) 10 kΩ; (b) V to I converter, $R_1 = 1$ kΩ, 10 kΩ; (c) 5.7 V, 6.0V.
8. 100 MΩ, 0.1 Ω.
12. $v_o = -(R/L)\int v_i\,dt$; integration.
15. For $R_1 = 10$ kΩ, $R = 200$ kΩ, $C = 795$ pF.
18. For $R_i \gg (R_F + R_o)$ and $R_F \gg R_o$, $R_{iF} = (R_F + R_o)/(A - R_o/R_F) = 0.011\,\Omega$.
22. 575 or 55.2 dB.
26. (a) $f_2 = 10$ Hz; (b) $f_2 = 10$ kHz.
31. (a) $d^2y/dt^2 - ky = 0$; (b) $y(t) = A \cos \sqrt{k}\,t + B \sin \sqrt{k}\,t$.

CHAPTER 21

1. $Qu = lANqu = l \cdot dQ/dt = li$.
4. $I \cong 4$ A.
6. (a) $H = 796{,}000$ A/m, $M = 16$ A/m; (b) $H = 318$ A/m, $M = 796{,}000$ A/m.
9. At $D = 1$ mm, $B = 2$ mT; at $D = 1$ cm, $B = 0.2$ mT.
11. Ni = 320 A-t.
14. (a) 300; (b) 1050; (c) 5300; (d) 4200.
16. 850 A.
18. (a) 0.7 A; (b) 4.2 A; (c) impossible.
22. 12.3 A.

CHAPTER 22

1. 12,000 V.
3. 1.05 mH, 2.1 H.
6. (a) 312 t; (b) 0.163 A.
9. $N' = 1.18\,N_o$.
11. (a) 1250, 15 t; (b) 0.6 A.
13. 30 A.
16. $a = 4$, $G = 0.5$ m$\mho$, $B = 2$ m$\mho$, $R = 2\,\Omega$, $X = 6\,\Omega$.
18. (a) $253\angle\underline{1.7^\circ}$ V; (b) $20.1\angle\underline{-1.4^\circ}$ A.
21. (b) $2.4\angle\underline{-90^\circ}$ A, $80\angle\underline{-39^\circ}$ A; (c) 97.8%.
23. $a = 1/20$, 20 dB.
26. (b) 21.8 A, 43.6 A.
29. 94.6%.
32. 1%, 98.5%.

CHAPTER 23

2. (a) $\mu_o I/2\pi D$, $I/2\pi D$; (b) $eU\mu_o\,I/2\pi D$; (c) $\mu_o\,I\,I_2\,dl/2\pi D$; (d) $\mu_o\,I\,I_3\,dl/2\pi D$; (e) 0.
5. (a) $\beta\,l\,u^2t/R$; (b) $\beta^2\,l^2\,u^3t^2/R$.
8. 955 A-t, 30 t.
11. 0.03015 Ω, 9994 Ω.
14. (a) $(NBlrn\pi/15) \cos (n\pi/30)t$; (b) $\mathrm{T}_m/2NBlr$, $\omega \mathrm{T}_m \cos \omega t$.
18. (a) 60 Hz; (b) 252 V.
21. 3055 rpm.

24. 0.004 N-m.
29. 10 A.

CHAPTER 24

1. (b) 270 V; (c) 223 V.
4. 295 V, 1250 W.
6. (a) 1.1 A; (b) 1.3 A.
9. 275 V, 45 V.
13. (a) 1558 rpm; (b) 1664 rpm; (c) 2400 rpm.
15. (a) I_A same, $n/2$; (b) $I_A \times 2$, $n/4$; (c) $I_A/\sqrt{2}$, $n/\sqrt{2}$.
17. $R_1 = 2.625\ \Omega$.
20. (a) 1300 rpm; (b) 1950 rpm; (c) 83.3 A, 122 N-m, 680 rpm.
23. (a) 1.3 A; (b) −80 A, 163 N-m.
27. 89%.

CHAPTER 25

1. (b) 1500 rpm.
2. (c) 4.
4. (a) 12 poles, 13,600 A; (b) $21{,}000\underline{/33.4^\circ}$.
6. $X_s = 0.5\ \Omega$.
9. I must be lagging V.
12. (b) $n_2 = n_1$, $\delta_2 \cong \frac{1}{2}\delta_1$, $\theta_2 > \theta_1$, $I_2 \cong \frac{1}{2} I_1$.
15. For $I_F = 1.5$ A, $I \cong 39$ A; $I_F = 3$ A, $I \cong 55$ A.
17. Old Transformer ok; new pf = 0.83.
20. 2 V, 0.067 V.
23. (a) $V_A = 240$ V; (b) $B_0 = -0.2\ \mho$, $G_0 = 0.018\ \mho$; (c) 82 A, 218 N-m, 54 hp.
26. 165 N-m.
29. (a) 20.9 kW or 28 hp; (b) 1197 rpm.
31. $n_2 = 900$ rpm, $T_2 = T_1$, $P_2 = 75$ hp.

CHAPTER 26

2. (b) Antenna direction, potentiometer, voltage difference detector, input direction.
4. (a) Pole at $s = -2\,\mathrm{a}$, zero at $s = -\mathrm{a}$; (b) $y(t) = -\frac{1}{2} + \frac{3}{2} e^{-2at}$.
7. (a) $10/(s + 2)$; (b) $6\,s/(s + 0.02)$.
10. (a) $T(s) = \infty$ means output with no input; (b) $y_n(t) = A_1 + A_2 e^{-bt}$.
14. (a) $T(j\omega) = 1/(0.1 + j\omega)$; (b) $y(t) = 10 - 10\,e^{-t/10}$.
15. (a) 100 V; (b) $V(t) = 84.5 + 82.5\,e^{-5t}$ V.

CHAPTER 27

2. Shunts: 1111 Ω, 101 Ω, 10.01 Ω.
4. Multipliers: 20, 90, 490, 1490 kΩ.
7. (a) 51.1 mA; (b) 50.44 mA.
10. 1.46, 1.11, 1.41 V.
14. 0.762 V.
16. (a) 360 Ω; (b) 243 Ω.
18. $R_X = 160\ \Omega$, $L_X = 6$ mH.
20. 4×10^{-4} in./in., 10 Hz.

Index

AC circuit, 123, 169
Accuracy of a measurement, 714
Active element, 22
Active network, 211
Acceptor atom, 294
Adder circuit, 394, 396
Admittance, 106, 123
Air gap, 558
Algebra, complex, 71
  block diagram, 235
  Boolean, 390
Alpha factor, 343
Alternating current, 67
Alternator, 612, 655
  construction of, 612
  model of, 656
  performance of, 657
Ammeter, 710
  connection of, 606
  construction of, 80
  multirange, 711
  shunts for, 606
Ampère, André, 551
Amplification, 347
Amplifier, 273
  cascading, 481
  class B and C, 431
  classification of, 409, 431
  design considerations in, 412, 427
  differential, 509
  elementary, 278, 347
  feedback, 484
  large-signal, 426
  multistage untuned, 481
  operational, 503
  power, 423
  power and efficiency in, 428
  practical, 412
  push-pull, 432
  RC-coupled, 465
  small-signal, 465
  tuned, 433
Amplifier impedance, 467, 490
Amplifier performance, 281
Amplifier response, 410, 466, 469
  low-frequency, 471
  high-frequency, 476
Amplitude distortion, 410
Amplitude modulation, 441
Analog computer, 523
Analogous equations, 89
Analogs, 89, 134
  electric and magnetic, 558
  table of, 135
Analog-to-digital converter, 732
Analysis, graphical, 53
  load-line, 54
  logic circuit, 392
  power series, 52
AND gate, 366
Apparent power, 174
Armature, 558, 613, 632
Armature reaction, 638
Audio amplifier, 410

Automatic control system, 685
Autotransformer, 595
Avalanche effect, 304, 424
Average value, 77
Avogadro's law, 289

Bandwidth, 190, 466
Bardeen, John, 255, 340
Barnes, John, 241
Base-spreading resistance, 454
Beam power tube, 278
Bell, Alexander Graham, 246
Belove, Charles, 434
Beta cutoff frequency, 374
Beta factor, 345, 453
Bias design, approximate, 422
Biasing, 301, 412, 425
  BJT, 417
  FET, 414
Bias stabilization, 487
Bilateral energy conversion, 599, 633
Binary arithmetic, 389
Binary counter, 380
Binary number system, 387
Binary-to-decimal conversion, 388
Bipolar junction transistor, 340
  biasing, 417
  model, 450, 469, 476, 479
  parameters, 458
  small-signal model, 453
Bistable multivibrator, 377
Black, H. S., 232
Black, L. J., 657, 678
Blanking circuit, 266
Block diagram, 234, 686
  of a voltage regulator, 698
Block diagram algebra, 235
Blocked-rotor test, 676
Bode chart, 246
Bode, H. W., 246
Bohr model, 267
Boltzmann constant, 268
Boole, George, 390
Boolean algebra, 390
Branch of a circuit, 29
Brattain, Walter, 255, 340
Breakdown phenomena, 304, 337
Breakdown torque, 673
Breakover voltage, 355
Breakpoint frequency, 247, 482
Bridge, Maxwell, 726
  impedance, 726
  Wheatstone, 725
Bridge circuit, 216
Bridge measurement, 724
Bridge rectifier circuit, 316
Bridged-T filter, 179
Brown, Gordon, 694
Bush, Vannevar, 576
Bypass capacitor, 472

Calibrating equipment, 732
Capacitance, 12, 19
  diffusion, 413
  interelectrode, 277
  junction, 413
  wiring, 413
Capacitance sensor, 724
Capacitive reactance, 124
Capacitor, 12
  by-pass, 472
  coupling, 471
Capacitor filter, 318
Capacitor motor, 678
  approximate analysis, 319
Carnot efficiency, 538
Carrier, charge, 295, 299
Carrier wave, 441
Cascading, 481
Cathode construction, 268
Cathode-ray oscilloscope, 264, 719
Cathode-ray tube, 261
Channel, communication, 442
Characteristic equation, 94
Charge, definition, 8
Checking, dimensional, 15
Checking circuit solutions, 37
Chemical energy, 536
Circuit, 4, 16
  ac steady-state, 169
  coupling, 220
  first-order, 150
  integrated, 348
  magnetic, 558
  mechanical, 41
  nonlinear, 51
  parallel, 16, 130
  second-order, 93, 154
  series, 16
  switching, 365
  three-phase, 192
  waveshaping, 325
Circuit analysis, 128, 132
Circuit element, 18

Circuit laws, 29
Circuit model, 5, 312, 439
  small-signal, 448, 453
  transformer, 580, 586
Clamping circuit, 328
Clark, R. N., 705, 706
Class A, B, and C operation, 431
Clipping circuit, 327
Closed-loop system, 687
CMOS units, 352, 385
Coercive force, 574
Collector cutoff current, 343
Color code for resistors, 748
Combinational circuit, 376
Common-base operation, 344
Common-base model, 450
Common-collector amplifier, 490
Common-mode rejection, 520
Common-mode voltage, 509
Communication system, 458
Common-emitter operation, 345
Common-emitter model, 452
Commutating pole, 639
Commutator, 613, 632
Comparator, op amp, 515
Compensating winding, 639
Compensation theorem, 711
Complete response, 148
Complex admittance, 124
Complex algebra, 61
Complex conjugate, 74
Complex impedance, 122
Complex numbers, 69
Complex plane, 68, 98
  roots in, 98
Complex power, 173
Compliance, 20, 135
Component, circuit, 16
  nonlinear, 51
Compound generator, 637
Compound motor, 645
Computer, analog, 523
Computer organization, 403
Conductance, 11, 125, 289
Conduction, 11
  in metals, 288
  in semiconductors, 291
  in solids, 287
Conductivity, 289
Connection equation, 33
Conservation, of charge, 8, 31
  of energy, 31, 538
Contact potential, 300
Continuity of stored energy, 21
Controlled source, 45, 448, 453
Control system, feedback, 685, 689
Conversion factors, 739
Converter, voltage-current, 514
Conwell, Esther, 291
Copper loss, 578, 585
Core loss, 578, 584
Core memory, 404
Corner frequency, 247
Coulomb, 8
Counter, binary, 380
  electronic, 718
Coupling, amplifier, 412
  capacitor, 471
  circuit, 220
  conductive, 221
  inductive, 222
  transformer, 224, 426
Covalent binding, 280, 536
Cramer's rule, 35
Critical damping, 98
Crystal, metallic, 288
Crystal of silicon, 290
Current, definition, 8
Current divider, 43, 210
Current source, 23, 420
Cutoff, Triode, 276
Cutoff frequency, 466, 472
Cutoff region, 371
Cyclotron, 284

d'Alembert's principle, 41, 134
Damped sinusoid, 97
Damping coefficient, 97
Damping ratio, 693
d'Arsonval movement, 605, 710
  circuit model of, 607
  step response of, 693
Dash pot, 21, 135
D'Azzo, J. J., 706
Decibels, 246, 482
Decimal notation, 14
Decimal-to-binary conversion, 388
Decision elements, 376
Definitions of electrical quantities, 7
Deflection system, CRT, 262
deForest, Lee, 255, 274
DeMorgan, Augustus, 392
De Morgan's theorem, 392
Delta connection, 194, 196

Depletion, 299
Depletion-mode FET, 338
Design, amplifier, 427
  BJT, 422
  JFET bias, 416
  logic circuit, 397
Detector, 443
Determinants in analysis, 34
Device, definition, 4
Differential amplifier, 509, 519
Differential transformer, 730
Differentiating circuit, 326
Differentiator, op amp, 511
Diffusion, 296
Digital computer, 403
Digital device, 363
Digital voltmeter, 719
Dimensional checking, 15
Diode, 297, 443
  four-layer, 354
  gas, 272
  light-emitting, 306
  models, 311
  photo, 306
  photo-emissive, 272
  Schottke, 306
  semiconductor, 297
  special purpose, 303
  tunnel, 305
  vacuum, 267, 312
Diode characteristics, 271, 299
Diode circuit model, 444
Diode logic circuit, 365, 370
Diode switch, 370
Direct current, 61
  response to, 119
Direct-current machines, 632
Discriminant, 94
Distortion in amplifiers, 410
  reducing, 488
Diverter, 635
Domain, magnetic, 552
Donor atom, 294
Doped semiconductors, 293
Double-dabble process, 388
Drift, 12, 288
Dual circuit construction, 139
Dual quantities, table of, 137
Duals, 136, 186
  table of, 137
Duodecimal system, 387
Dynamic junction resistance, 444, 451
Dynamic pickup, 603
Dynamic resistance, 443
  FET, 446
Dynamic response, 689
Dynamo, elementary, 608

ECL gate, 386
Eddy currents, 575, 579, 633
Effective value, 78, 170
Efficiency, amplifier, 428
  Carnot, 538
  generator, 620
  transformer, 589
Einstein equation, 297
Electric field, motion in, 256
Electric field focusing, 262
Electric field strength, 8, 9, 257
Electrical energy, 538
Electrical engineering, 3
Electrical machines, 631
Electrodynamometer instrument, 715
Electromagnet, 621
Electromechanics, 597
Electromotive force, 608
Electron, model of, 256
Electron deflection, 262
Electron emission, 267
Electron gun, 10, 262
Electron microscope, 261
Electronic instruments, 716
Electronic switching, 364, 369
Electronic voltmeter, 716
Electronics, history of, 255
Electron motion, 256, 288
  in E and B fields, 261
Electron tube, future of, 280
Electron volt, 258
Electrostriction, 599
Element equation, 33
Element, circuit, 16
  linear, 18
  nonlinear, 18, 51
  symbols for, 25, 601
  unintentional, 413
Emf, generated, 608, 632
  induced, 578
Emission, electron, 267, 268
Emitter by-pass calculations, 474
Emitter follower, 491
Energy, 3, 8
Energy electrical, 10, 538
Energy balance, 619

Energy bands, 267
Energy conversion, 539
  bilateral, 599
  electrical, 538
Energy density, magnetic, 573
Energy dissipation, 20
Energy gained by electron, 257
Energy source, 22, 536
Energy storage, 19
  magnetic field, 572
Energy supply, 535
Enhancement-mode FET, 337
Enke, C. G., 713
Entropy, 537
Epitaxial growth, 298
Equality comparator, 395
Equivalent circuit, 42, 209
Error, instrument, 714
Esaki, Leo, 305
Esaki diode, 305
Euler's theorem, 69
Evans, W. R., 704
Excitation, generator, 634
Exciting current, transformer, 579
  induction motor, 671
Exclusive-OR gate, 395
Experimental laws, 11
Exponential function, 62, 65
  response to, 118
Exponential solution, 89

Fan-in, fan-out, 370
Faraday, Michael, 13
Faraday's law, 570
Feedback, 232, 484, 686
  negative, 238, 486
  positive, 236, 484
Feedback amplifier, 233, 484
Feedback control system, 698
Feedback system, basic, 235, 689
Ferromagnetism, 552, 574
Field, 3, 9
  due to currents, 551
  electric, 8, 256
  magnetic, 8, 549
  rotating, 659
Field around a conductor, 553
Field connections, 635
Field effect transistor, 335
  biasing, 414
  model, 448, 469, 476
Field excitation of machines, 634
Filter, 177, 221, 318
  Bridged-T, 179
  capacitor, 318
  high-pass and low-pass, 178, 512
  L-section, 321
  power supply, 318
First-order circuit, 150
First-order system, 88, 243
Fission, nuclear, 537
Flip-flop, 377, 380
Fluorescent screen, 264
Flux linkage, 570
Flyball governor, Watt's, 685
Focusing, electric field, 262
Follower, 490, 687
  emitter, 491
  source, 492
  voltage, 507
Forbidden bands, 267
Force, 3, 7
  developed, 599
Force on electron, 257, 259
Forced response, 87, 117
Forcing function, 81
  response to, 117
Formulation of equations, 39
Forward bias, 301
Fourier analysis, 66, 410
Four-layer diode, 354
Frank, Ernest, 714
Frequency, natural, 97, 693
  resonant, 180
Frequency comparison, 266
Frequency distortion, 411
Frequency divider, 380
Frequency modulation, 442
Frequency response, 177, 182
  amplifier, 411, 433, 466, 469
  improving, 245, 486
  in decibels, 482
  of a system, 697
  op amp, 522
Frequency selectivity, 182
Friction, 20
Fringing of magnetic flux, 559
Fuel cell, 540
Full-adder, 396, 399
Full-wave rectifier, 316
Function generator, 524, 528
Fusion, nuclear, 537

Gain, 233, 410
  in decibels, 482

  stabilization of, 238
Gain-bandwidth product, 482
Gardner, M. F., 241
Gas diode, 272
Gate, logic, 364
Gauss' law, 270
Generator, 22, 600
  alternating current, 656
  compound, 637
  construction of, 614, 632
  direct current, 632
  elementary, 193
  model of, 633
  self-excited, 635
  shunt, 635
Geothermal energy, 537
Germanium, properties of, 290
Giacoletto, L. J., 455
Graphical analysis, 53, 279
Gray, P. E., 382
Grid, vacuum-tube, 275
Guidance system, 231

Halacsy-von Fuchs, 569
Half-adder circuit, 394
Half-power frequency, 178, 190, 466, 472
Half-wave rectifier, 315
Halkias, C., 303, 318, 410, 434
Hall effect, 310
Ham, J. M., 552
Harris, J. N., 382
Henry, Joseph, 13
Hibberd, Robert, 359, 382
High-field emission, 268
High-frequency model, 449, 455, 476
High-frequency response, 476
High-pass filter, 471
History of electronics, 255
Hoadley, G. B., 576
Holes and electrons, 290
Homogeneous equation, 89
Horse-power rating, 80
Houpis, C. H., 706
h-parameter model, 455
Hybrid parameters, 455, 583
Hybrid-π model, 454, 479
Hysteresis loss, 574, 579, 633

Ideal transducer, 273, 600
Ideal transformer, 587
Ignition coil, 22
Imaginary number, 68
Immittance, 131
Impedance, 105, 123
  amplifier, 413, 467
  complex, 122
  input and output, 413, 490
  motional, 134
  transfer, 216
  to exponentials, 100
Impedance bridge, 726
Impedance function, 101, 105
Impedance matching, 212
Impedance transformation, 428, 492
Impedance triangle, 174
Impulse function, 161
Impulse response, 162
Induced voltage, 570
Inductance, 13, 19, 571
  self and mutual, 13, 221
Induction motor, 667
  model of, 671
  performance of, 676
  single-phase, 677
Inductive reactance, 124
Inertia, 21
Information processing, 3, 363
Initial conditions, 90, 156
Injection current, 302
Input-output computer, 404
Instantaneous value, 79
Instrument, 710
  electrodynamometer, 715
  electronic, 716
  indicating, 710
  iron-vane, 617
  recording, 710
  rectifier, 716
Instrumentation, 709
Instrumentation system, 731
  automatic, 734
Instrument errors, 714
Insulator, 287
Integrated circuit, 348
  fabrication, 350
Integrating circuit, 327
Integrator, op amp, 511
Interelectrode capacitances, 277
Intrinsic concentration, 292, 295
Inverter, 367, 376, 379
Inverting amplifier, 504
Iron-vane instrument, 617
Iron loss, 578
Irreversible transformation, 20

JFET biasing, 415
Joule, definition, 8
Junction capacitance, 455
Junction diode, 299
Junction formation, 298
Junction laser, 306
Junction transistor, 340

Karnaugh, Maurice, 400
Karnaugh map, 400
Kilby, J. S., 255
Kirchhoff, Gustav, 29
Kirchhoff's laws, 30, 31, 123
Klystrons, 280

Laplace transformation, 241
Large-signal amplifier, 426
Lattice, 288
Laws, experimental, 11
  of thermodynamics, 537
Leach, Donald P., 382
Leakage flux, 560, 580
Lenz's law, 571
Light-emitting diode, 306
Linearity, 17, 48, 89, 210, 443
Linearization, piecewise, 53
Linear model, 443
Lissajous pattern, 266
Load impedance, 412
Load line, 54, 279
  dc, 426
  location of, 425
Logic circuit analysis, 392
Logic circuit design, 397
Logic circuit synthesis, 393
Logic classification, 369
Logic gate, diode, 366
  I.C., 402
  MOS, 375
  transistor, 372
Logic operation, 365
Loop of a circuit, 29
Loop-current method, 36
Losses, fixed and variable, 212
  machine, 633
  transformer, 575
Loudspeaker, dynamic, 602
Low-frequency response, 471
Low-pass filter, 477, 512
L-section filter, 321
LSI chips, 354
Lynch, W. A., 600

Machines, alternating-current, 655
  classification of, 631
  direct-current, 631
  induction, 667
  rotating, 612
  speed control of, 645, 678
  starting, 643, 678
  synchronous, 660
Magnetic circuit, 557, 563
  ac excitation of, 570
Magnetic circuit calculation, 558
Magnetic domain, 552
Magnetic field, 14, 549
  energy storage in, 572
Magnetic field intensity, 551
Magnetic field relations summary, 556
Magnetic flux density, 10, 550
Magnetic polarization, 553
Magnetic saturation, 553
Magnetizing current, 579
Magnetohydrodynamics, 545
Magnetization curve, 552, 560
  machine, 636, 657
Magnetomotive force, 554
Magnetostriction, 597
Magnetrons, 280
Majority carrier, 299
Malmstadt, H. V., 713
Malvino, A. P., 382
Mapping, Karnaugh, 400
Marconi, Guglielmo, 255
Mass, 19
  of an electron, 256
Mass-action law, 295
Mass spectrograph, 284
Mathematical model, 440
Maximum power transfer, 212, 428
Maxwell bridge, 726
Measurement, ac, 715
  bridge, 724
  dc, 710
  phase-difference, 266
  potentiometer, 727
  precision of, 714
  time, CRO, 265
  voltage, CRO, 265
Mechanical circuit, 41
Mechtley, E. A., 739
Memory, computer, 404
Memory device, 376
Metallic conductor, 288
Meter reading, 79, 711

Midfrequency gain, 469
Midfrequency range, 466
Miller, J. M., 476
Miller capacitance, 476
Millman, Jacob, 303, 318, 410, 434
Minority carrier, 300
Mmf law, 555
Mobility, 288
Model, 5, 17, 439
  alternator, 656
  amplifier, 467
  atomic, 267
  BJT, 450
  communication transformer, 588
  commutator machine, 615
  d'Arsonval movement, 607
  dc generator, 620, 633
  dc motor, 640
  diode, 311, 444
  FET, 448
  induction motor, 670, 673
  mathematical, 440
  synchronous motor, 662
  transducer, 601
  transformer, 580
Model, circuit, 312, 439
  h-parameter, 455
  high-frequency, 449, 476
  hybrid-π, 454
  small-signal, 439, 443
Modulation, amplitude, 441
  frequency, 442
Modulator, nonlinear, 443
Motional impedance, 134
MOSFET, 337
  biasing, 414
Motor, 22, 600
  capacitor, 678
  dc, 639
  induction, 667
  series, 642
  shaded-pole, 678
  shunt, 640
  split-phase, 678
  synchronous, 660
  wound-rotor, 676
Multimeter, 607, 714
Multiplexing system, 733
Multiplier resistance, 712
Multistage amplifier, 481
Mutual inductance, 13, 221
Mutual resistance, 37

NAND gate, 368
  DTL, 372
  synthesis, 398
  TTL, 374
Natural frequency, 97, 693
Natural response, 87
Negative feedback, 238, 486
Network, 16
  active, 44, 211
  linear, 49, 210
  non-linear, 51
  one-port, 42
  passive, 44, 209
  three-terminal, 216
  two-port, 216
  two-terminal, 124, 209
Network theorems, 42
Network reduction, 42, 200
Neutral connection, 200
Newton, definition, 7
Newton's law, 256
Node of a circuit, 29
Node-voltage method, 38
Noise, 410
No-load motor test, 676
Nomenclature, 119
  transistor, 348
  vacuum-tube, 280
Noninverting amplifier, 507
Nonlinear devices, 311
Nonlinear elements, 51, 53
Nonlinearity, 410, 440, 488
Nonlinear networks, 51
NOR gate, 367
Normalized response curves, 695
Norton's theorem, 45, 211
NOT gate, 367
n-type semiconductor, 293
Nuclear energy, 537
Nyquist stability criterion, 689

Offset current, 520
Offset voltage, 520
Ohm, Georg, 11
Ohmic contact, 306
Ohmmeter, 713
Ohm's law, 11, 289
Oliver, Barney, 735
One-port network, 209
Open circuit, 24
Open-circuit test, 584
Open-loop, closed-loop system, 687

Operating region, permissible, 424
Operational amplifier, 503, 744
  application, 508
  ideal, 504
  practical, 518
Operator *j*, 68
OR gate, 366
Oscillation, 237
Oscillator, 234, 458, 484
  sawtooth, 356
Oscillatory response, 97
Oscilloscope, 719
Output impedance, 413, 490

Parameter, lumped, 4
  transformer, 582
  transistor, 448, 458
Peak inverse voltage, 317
Peak value, 77
Pearson Gerald, 255, 304
Peatman, John, 382
Peltier effect, 543
Pentavalent element, 293
Pentode, 277
Performance, alternator, 657
  amplifier, 428
  dc motor, 647
  induction motor, 672, 676
  synchronous motor, 666
Period, 66
Periodic waveform, 76
Periodic table, semi-conductors, 293
Permeability, 551
Permissible operating region, 424
Permittivity, 270, 353
Phase angle, 66, 128
Phase inverter, 317
Phase rotation, 194
Phase-difference measurement, 266
Phase-shift oscillator, 485
Phasor, definition, 70
  effective value, 81
Phasor diagram, 75, 128, 182
Phasor method, 120, 122
Phillips, H. B., 98
Phosphorescence, 264
Photoconductive cell, 723
Photodiode, 306
Photoelectric emission, 268
Photoelectric transducer, 722
Photoemission diode, 272
Photoresist, 351
Phototube, 272
Physical constants, table of, 739
Pickoff point, 234
Pickup, dynamic phonograph, 604
  reluctance, 617
Piecewise linearization, 53
Piezoelectric effect, 597, 722
Pinch-off voltage, 337
π-T transformation, 218
Planck constant, 267
Planar technology, 298
pn junction, behavior, 297
  formation, 298
Point-contact transistor, 340
Polar form of complex number, 69
Poles and zeros, 102
  analysis of systems by, 696
  diagram of, 104
  general procedure for, 107
  physical interpretation of, 104
Polyphase circuit, 192
Positive feedback, 236, 484
Potential barrier, 300, 342
Potential difference, 9
Potential hill, 300
Potentiometer, 727
Power, 8, 10, 170
  active, 171
  apparent, 174
  complex, 173
  electrical, 10
  maximum, 212
  reactive, 170
Power amplifier, 423, 428
Power angle, 611
Power calculation, 170
  three-phase, 199
Power factor, 172
  correction of, 176, 666
Power-flow diagram, 647, 661
Power-series analysis, 52, 440
Power supply, 315
Power supply design, 321
Power transfer, maximum, 212
Power triangle, 174
Precision, instrument, 714
Procedure for circuit analysis, 132
  for natural behavior, 92
  for complete response, 148
  for using poles and zeros, 107
p-type semi-conductor, 293
Pull-out torque, 664

Pulse, 61, 160
Pulse response, 159
Push-pull amplifier, 432

Quality factor *Q*, 183
  basic definition, 187
Quantities, analogous, 135
  basic, 7
  derived, 8
  electrical, 9
Quiescent point, 280, 347, 425

Radar, 325
Rating, root-mean-square, 80
  transformer, 589
RC-coupled amplifier, 465
RC-tuned oscillator, 485
Reactance, 124, 171
Reactive power, 170
  in synchronous machines, 664
Reciprocity theorem, 215
Recombination, 294
Rectangular form of complex number, 69
Rectifier, 315
  silicon controlled, 356
Rectifier instrument, 715
Register, serial shift, 381
Regulation, voltage, 590
  speed, 642
Regulator, 324, 687
Relay, 364, 616, 622
Reluctance, 557
Reluctance pickup, 617
Reluctance torque, 618, 664
Residual magnetism, 574
Resistance, 11, 20, 289
  alternating-current, 124
  base-spreading, 454
  dynamic, 443
  frictional, 20, 134
  in metallic conductor, 289
  preferred values, 748
Resistance sensor, 723
Resistivity, 289
Resistor, electrical, 11
  color code for, 748
  nonlinear, 51
Resonance, 180
  parallel, 185
  series, 181
Resonant frequency, 180, 186
Response, complete, 148
  dynamic, 243, 689
  forced, 87, 117
  frequency, 177, 181
  impulse, 162
  instrument, 715
  natural, 87
  pulse, 159
  step, 154
  to direct-current, 119
  to exponentials, 118
  to sinusoids, 120
Reverse current, 301
Reverse bias, 301
Reverse saturation current, 302
Reversible transformation, 22, 602
Rheostat, 635
Richardson-Dushman equation, 284
Right-hand rule, 260
Ripple, commutator, 614
Ripple factor, 317
Rms value, 170
Robertson, B. L., 657, 678
Root locus, 99, 704
Root-mean-square rating, 78, 80
Roots of characteristic equation, 94
Rotating machine, 612
Rotating magnetic field, 659
Rotational transducer, 684

Saturation current, 302
Saturation region, 372
Sawtooth wave, 62
Schilling, D. L., 434
Schottky diode, 306
Screen grid, 277
Searle, C. L., 382
Second harmonic, 411
Secondary emission, 268, 277
Second-order circuit, 154
Second-order system, 93, 244, 690
Seebeck effect, 542
Self-biasing circuit, 419
Self-excited generator, 635
Self inductance, 13
Self-induction, 13, 221
Semiconductor, 288
Semiconductor diode, 297, 313
Sensor, 723, 731
Sequential circuits, 377
Series motor, 642
Servomechanism, 687
Sevin, Leonce, 359

Shaded-pole motor, 678
Shannon, Claude, 390
Sheet resistance, 352
Shockley, William, 255, 340
Short circuit, 24
Short-circuit test, 584
Shot effect, 411
Shunt generator, 635
Shunt motor, 640
Shunt resistance, 710
Sideband, 442
Signal degradation, 370
Signal generator, 458, 515
Signal size, effect of, 443
Signals, 62
  large and small, 439
Silicon, properties of, 290
Silicon controlled rectifier, 356
Single-phase induction motor, 677
Sinusoid, 62, 66
  response to, 120
Skilling, H. H., 85, 71, 619, 678, 711
Slemon, G. R., 552
Slew rate, 523
Slip, 668
Small signals, 439
Small-signal model, 467
  BJT, 451, 453
  FET, 448
Small-signal amplifier, 465
Solar cell, 541
Solar energy, 536
Solenoid, 550, 555
Source, 23
Source, controlled, 45, 448, 453
Source follower, 492
Space charge, 268
Speed control, 643, 645, 674, 678
Speed regulation, 642
Split-phase motor, 678
Spring characteristic, 20
Square-wave generator, 516
Squirrel-cage motor, 668, 676
Stability, 688, 702
Stabilizing operation, 486
Star connection, 195
Steady-state condition, 87, 169
Steinmetz, Charles, 68
Step response, 61, 154, 240, 701
Stern, Lothar, 359
Strain-gauge, 729
Substitution method, 35
Substrate crystal, 298
Summing amplifier, 506
Summing point, 234
Superposition theorem, 48, 210
Suppressor grid, 278
Surface barrier energy, 268
Susceptance, 125
Sweep generator, 265
Switch, diode, 370
  transistor, 371
Switching, electronic, 365
Switching logic, 364
Switching transient, 151, 157
Symbols, 25, 601
Synchronous condenser, 665
Synchronous machine, 655
Synchronous motor, 660
  performance of, 666
Synchronous reactance, 656
Synchronous speed, 656
Synthesis, logic circuit, 393
  NAND-gate, 398
System, 5, 231, 685
  automatic control, 685
  communication, 458
  feedback control, 689
  first-order, 88, 243, 690
  instrumentation, 731
  open- and closed-loop, 687
  oscillatory, 97
  second-order, 93, 244, 690, 695
  underdamped, 97
System engineering, 231
System response, 688, 695

Tachometer, 723
Tape recorder, 733
Telemetry, 733
Terman, F. E., 281
Tesla, Nicola, 667
Tesla, definition, 10
Tetrode, 277
Thermal analogs, 135
Thermal generation rate, 294, 302
Thermal runaway, 418
Thermal velocity, 289
Thermionic converter, 544
Thermionic emission, 268
Thermistor, 292
Thermocouple, 542, 715
Thermodynamics, laws of, 537
Thermoelectric converter, 542
Thermoelectric transducer, 722

Thévenin's theorem, 44, 211
Three-phase circuit, 192
Threshold voltage, 304, 372
Thyristors, 354
  control, 646
Timbie, W. H., 576
Time constant, 63, 65, 92, 245
  system, 695
Time measurement, 265
Toren, E. C., Jr., 713
Toroid, 555
Torque, developed, 609, 633, 663
  reluctance, 619, 664
Torque angle, 611, 661
T-π transformation, 218
Transadmittance, FET, 447
Transconductance, 446, 454
Transducer, 4, 459, 722, 731
  dynamic, 602
  ideal, 273, 600
  instrumentation, 722
  moving-iron, 616
  rotational, 604
  translational, 598
Transfer characteristics, 276, 346
  FET, 338
Transfer function, 234, 240, 690
Transfer impedance, 216
Transform, 71, 76, 136, 241
Transformation, irreversible, 20
  reversible, 22, 602
Transformer, 224, 569
  circuit model, 580, 587
  construction, 577
  coupling, 426
  current relation, 579
  differential, 730
  efficiency, 589
  exciting current, 579
  frequency response, 588
  ideal, 587
  impedance, 428, 492
  iron core, 569
  model, 588
  nameplate rating, 589
  operation, 576
  performance, 585, 589
  voltage regulation, 590
  voltage relation, 578
Transient period, 87
Transient, switching, 151, 157
Transit time, 280
Transistor, 335, 340
  biasing, 415
  carrier motion in, 343
  characteristics, 337, 344, 740
  circuit model, 373, 448
  fabrication, 341
  field-effect, 335
  junction, 340
  point-contact, 340
  typical parameters, 458
Transistor logic circuit, 367, 372
Transistor nomenclature, 348
Transistor switch, 371
Transition region, 299
Traveling-wave tube, 280
Triangular-wave generator, 517
Trigger circuit, 265
Triode, characteristics of, 276, 746
Triode voltage amplifier, 278
Trivalent element, 293
Truth table, 365
Truxal, J. G., 600, 706
Tube, cathode ray, 261
Tube characteristics, 746
Tuned amplifier, 433
Tunnel diode, 305
Turn ratio, electromechanical, 600
  transformer, 578
Two-port, 42, 216, 273, 445, 600

Undamped natural frequency, 99
Underdamped response, 97
Units, SI, 7
Unity-gain buffer, 510

Vacuum diode, 267, 268, 312
  characteristics of, 270
Vacuum triode, 274
Vacuum-tube nomenclature, 280
Valence electrons, 267, 291
Vector representation, 16
Velocity, drift, 288
  electron, 257
Velocity of light, 259
V-I characteristic, 16
Video amplifier, 410
Virtual cathode, 270
Virtual work method, 617
Volt, definition, 9
Volt-ampere, 173
Voltage, definition, 9
  generated, 193, 608

  induced, 570
  measurement with CRO, 265
Voltage divider, 43, 210
Voltage follower, 507
Voltage gain, 280
Voltage generation, 193
Voltage generator, 23
Voltage gradient, 9
Voltage regulation, 590
Voltage regulator, 698
  Zener diode, 324
Voltage source, 23
Voltaic cell, 540
Voltmeter, 710
  connection of, 606
  digital, 719
  electronic, 716
  multiplier for, 606
Vote taker design, 397

Ward-Leonard system, 646
Watt, definition, 8
Wattmeter, 170, 199
Waveform, 61, 76
  display of, 265, 720
Waveshaping, 325
Weber, definition, 8
Wheatstone bridge, 725
Wide-band amplifier, 410
Wien bridge, 486
Wilson, V. C., 544
Work function, 268
Wound-rotor motor, 676
Wye connection, 195

X-Y plotting, 266

Y-Δ transformation, 218

Zener diode characteristics, 314
Zener diode voltage regulator, 324
Zener effect, 304
Zeros and poles, 104

LIST OF SYMBOLS

a	acceleration, turn ratio
A	amplifier gain, area
A	amperes
B	magnetic flux density, susceptance
BW	bandwidth
C	capacitance
C	coulombs
C_p	thermal capacity
d	distance
dB	decibel
D	diffusion coefficient, electric flux density, frictional resistance
e	charge of an electron,
e, E	electromotive force
e	natural log base 2.718 . . .
$\mathcal{E}$	electric field strength
f	frequency
f, F	force
F	farads
$\mathfrak{F}$	magnetomotive force
g	thermal generation rate
g_m	transconductance
G	conductance
$\mathbf{G}$	transfer function
h	hybrid parameter, Planck's constant
h	hours
H	magnetic field intensity
H	henrys
$\mathbf{H}$	feedback function
i, I	current
IC	integrated circuit
j	complex operator $\sqrt{-1}$
J	current density, rotational inertia
J	joules
k	Boltzmann's constant
k, K	constant of proportionality
K	compliance
K	degrees Kelvin
l	length
L	inductance
m	mass of an electron
m	meters
M	mass, mutual inductance
n	electron concentration, rotational speed
n, N	number of elements
N	newtons
p	hole concentration, number of poles
p, P	power
q, Q	electric charge
q_T	rate of heat flow
Q	quality factor
r	radius, ripple factor, small-signal resistance
R	recombination rate, resistance
R_T	thermal resistance
$\mathcal{R}$	reluctance
s	complex frequency,

s	slip
s	seconds
t	time
T	period, time constant
T	teslas
T	transfer function
u, U	velocity
v, V	voltage
V	volts
w, W	energy
W	watts
W	complex number
Wb	webers
x, X	displacement
X	reactance
Y	admittance
Z	impedance

Also see:
Table 1-1 p. 7
Table 1-2 p. 8
Table 1-3 p. 15
Table 5-1 p. 135
Table 13-1 p. 348
Table 21-1 p. 558
Table 23-1 p. 601
Summary p. 382

α	(alpha) forward current transfer ratio (CB)
β	(beta) forward current transfer ratio (CE)
γ	(gamma) emitter efficiency
δ	(delta) torque angle, fractional detuning
Δ	(delta) incremental change
ϵ	(epsilon) permittivity, strain
ζ	(zeta) damping ratio
θ	(theta) angle
λ	(lambda) flux linkage, luminous flux
μ	(mu) mobility, permeability
π	(pi) constant 3.14159 . . .
ρ	(rho) charge density, resistivity
σ	(sigma) conductivity, real component of **W**
Σ	(sigma) summation
τ, T	(tau) torque, temperature, time constant
ϕ, Φ	(phi) magnetic flux
ψ	(psi) electric flux
ω, Ω	(omega) angular velocity
ω	frequency, imaginary component of **W**
Ω	ohms
℧	mhos

ISBN 0-471-80171-2